University of Alberta

REGULATION OF THE HUMAN DEAD BOX GENE, *DDX1*, IN RETINOBLASTOMA AND NEUROBLASTOMA CELL LINES.

By

Wenjun Bie ©

A thesis submitted to the faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Science.

Medical Sciences-Oncology

Edmonton, Alberta

Fall, 1996

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-612-18235-5

Canada

University of Alberta

Library Release Form

Name of Author: Wenjun Bie

Title of Thesis: Regulation of the human DEAD box gene, *DDX1*, in retinoblastoma and neuroblastoma cell lines.

Degree: Master of Science

Year this Degree Granted: Fall, 1996

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other right in association with the copyright in the thesis, and except as hereinbefore provided neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material from whatever without the author's prior written permission.

(Signed) Wenjun Bie

Permanent address:
Molecular Oncology Program
Cross Cancer Institute

Date: August 14, 1996

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled REGULATION OF THE HUMAN DEAD BOX GENE, *DDX1*, IN RETINOBLASTOMA AND NEUROBLASTOMA CELL LINES submitted by Wenjun Bie in partial fulfillment of the requirements for the degree of Master of Science in Medical Sciences -- Oncology.

Roseline Godbout

Dr. Roseline Godbout (supervisor)

Don W. Morrish

Dr. Don W. Morrish

Malcolm C. Paterson

Dr. Malcolm C. Paterson

Andrew R. E. Shaw

Dr. Andrew R. E. Shaw

Date: Aug. 12/96

DEDICATION

This thesis is dedicated to my parents who are always an inspiration to me, and to my husband and my son, whose encouragement and support gave me the courage to achieve my goals. My success is theirs.

ABSTRACT

DEAD box proteins are a family of proteins that are characterized by eight highly conserved amino acid motifs, one of which is the DEAD motif (DEAD standing for Asp-Glu-Ala-Asp). These proteins are putative RNA helicases that are believed to be involved in many cellular functions such as translation initiation, RNA splicing and ribosomal assembly. The human DEAD box gene, *DDX1*, was identified by differential screening of a RB-enriched cDNA library. *DDX1* was shown to be amplified and over-expressed in two retinoblastoma and several neuroblastoma cell lines amplified for *MYCN*, as well as in some primary neuroblastoma tumours amplified for *MYCN*, suggesting co-amplification of *DDX1* and *MYCN* in these cell lines and tumours.

To study the regulation of *DDX1* transcription in retinoblastoma and neuroblastoma cell lines, we have cloned the 5' end of the DDX1 cDNA. Three major and one minor transcription initiation sites of *DDX1* have been identified. DNase I hypersensitive sites, specific DNA-protein interactions and DNase I protected regions have been detected in the 5' flanking region, as well as in the first exon and first intron of *DDX1*, suggesting that these regions contain regulatory elements. In addition, one region in *DDX1* was identified which has homology to the *MYCN* promoter region.

ACKNOWLEDGMENT

I would like to express my special thanks to my supervisor, Dr. Roseline Godbout, for her supervision and support throughout my M. Sc. studies. I would also like to express my thanks to the members of my supervisory committee, Drs. Rufus S. Day III, Don W. Morrish and Malcolm C. Paterson for their time and advice.

Thanks to all people in the lab, Mary Packer, Elizabeth Monckton, Randy Andison, Dwayne Bisgrove and Stacey Hume who contributed to this thesis by providing an enjoyable working environment, excellent technical assistance and advice. A special thanks goes to Sachin Karyal for his help in computer work. Thanks to Viki Bjerkelund and Gina Kennedy for their friendly advice and help.

Financial support during my study was provided by the National Cancer Institute of Canada.

TABLE OF CONTENTS

Page

CHAPTER I. Introduction

I. Retinoblastoma and the *RB* gene 1

I. 1. General background 1

I. 2. The functions of RB protein (pRB) 2

I. 2. 1. The role of pRB in the cell cycle 2

I. 2. 2. The role of pRB in transcription 4

I. 2. 3. Are other genes involved in RB tumour formation? 6

II. The DEAD box protein family 7

II. 1. Conserved motifs in the DEAD box protein family 9

II. 2. The functions of the DEAD box proteins 10

II. 3. Regulation of DEAD box proteins 12

III. DDX1 13

IV. The aim of this study 14

CHAPTER II. Materials and Methods 16

CHAPTER III. Cloning the 5' end of DDX1 cDNA and identification of the transcription initiation site

III. 1. Introduction 25

III. 2. Results 26

Cloning the 5' end of DDX1 cDNA 26

Identification of the transcription initiation site 28

CHAPTER IV. Characterization of the regulatory elements of the *DDX1* gene

IV. 1. Introduction 34

IV. 2. Results 36

Genomic cloning 36

Mapping DNase I hypersensitive (HS) sites in *DDX1* 37

EMSA and DNase I footprinting assays 39

CHAPTER V. Discussion and future work

V. 1. Discussion 67

V. 2. Future work 73

REFERENCE 76

LIST OF FIGURES AND TABLE

Page

Figure 1. Schematic representation of highly conserved amino acid motifs in the DEAD box protein family 8

Figure 2. Schematic representation of amplification of the 5'-end of a gene using the 5'-AmpliFINDER RACE method 20

Figure 3. Sequence of the 5' end of the DDX1 2.7 kb cDNA obtained using the RACE strategy 27

Figure 4. Southern blot analysis of PCR products from the RACE experiments 29

Figure 5. Structural diagram of the most stable secondary structure at the 5' end of the DDX1 2.7 kb cDNA 30

Figure 6. Transcription start sites of *DDX1* determined by primer extension 32

Figure 7. Mapping of *DDX1* transcription initiation sites by primer extension 33

Figure 8. Sequence showing the *MYCN* promoter homology region in *DDX1* 38

Figure 9. Mapping of DNase I hypersensitive (HS) sites in a region from -2.1 kb to +2.1 kb of *DDX1* 40

Figure 10. Mapping of DNase I HS sites in a region from -1.1 kb to +4.1 kb of *DDX1* 42

Figure 11. Mapping of DNase I HS sites in the upstream flanking region of *DDX1* 44

Figure 12. Summary of DNase I HS sites of *DDX1* 46

Figure 13. Schematic representation of DNA fragments used in EMSA and DNase I footprinting 47

Figure 14. Analysis of the interaction between nuclear proteins and DNA fragment 1 (-577/-203) of *DDX1* by EMSA 49

Figure 15. Analysis of the interaction between nuclear proteins and DNA fragment 2 (-202/-36) of *DDX1* by EMSA 51

Figure 16. Analysis of the interaction between nuclear proteins and DNA fragment 3 (-35/+163) of *DDX1* by EMSA 54

Figure 17. Analysis of the interaction between nuclear proteins and DNA fragment 4 (+164/+431) of *DDX1* by EMSA 56

Figure 18. DNase I footprinting analysis of DNA fragment 1 (-577/-203) of *DDX1* 58

Figure 19. DNase I footprinting analysis of DNA fragment 2 (-202/-36) of *DDX1* 60

Figure 20. DNase I footprinting analysis of DNA fragment 3 (-35/+164) of *DDX1* 63

Figure 21. Sequence of DNase I protected regions I to VII 65

Table 1. Results of searching transcription factor binding sites for DNase I protected regions in *DDX1* 66

ABBREVIATIONS

CaM:	calmodulin
CAT:	chloramphenicol acetyl transferase
cDNA:	complementary deoxyribonucleic acid
Ci.	curie
cpm:	counts per minute
C-terminal:	protein carboxy terminal
DDX1:	DEAD box 1
DHFR:	dihydrofolate reductase
DMEM medium:	high glucose Dulbecco's Modified Eagle Medium
DNA:	deoxyribonucleic acid
DNase I:	deoxyribonuclease I
E. coli:	*Escherichia coli*
EDTA:	ethylenediaminetetraacetic acid, disodium salt
EMSA:	electrophoretic mobility shift assay
hnRNP U:	heterogeneous nuclear ribonucleoprotein U
kb:	kilo-base pairs
kDa:	kilo-Dalton(s)
min:	minute
mRNA:	messager RNA
μg:	microgram
N-terminal:	protein amino terminal
PCR:	polymerase chain reaction
p.f.u:	plaque forming unit
PKC:	protein kinase C
RACE:	rapid amplification of cDNA ends

RB:	retinoblastoma
RCE:	retinoblastoma control element
RNA:	ribonucleic acid
RNase:	ribonuclease
rpm:	revolutions per minute
SV40:	simian virus 40
TCE:	*TGF*-β control element
TLC:	thin layer chromatography
Tris:	tris (hydroxymethy) aminomethane
V:	volts

CHAPTER I.

INTRODUCTION

I. Retinoblastoma and the *RB* gene

I. 1. General background

Retinoblastoma (RB), a rare childhood tumour of the retina, occurs in both hereditary and nonhereditary forms. In the hereditary cases, children develop bilateral, mutifocal tumours of the retina at an early age. In contrast, nonhereditary RB is characterized by unilateral, unifocal tumours that develop in older children. In 1971, based on a statistical analysis of RB, Knudson postulated that RB could result from two mutations affecting both copies of a gene (1). He proposed that in the hereditary form, one mutation is inherited and a second mutation is acquired somatically, leading to tumour development. In the nonhereditary form, both mutations occur in somatic cells. Subsequent chromosomal abnormality, esterase D and DNA polymorphic analyses provided strong support for this hypothesis, indicating that these mutations are located on chromosome band 13q14 (2-5) and that both mutations involve the two alleles of the *RB* gene (6-10). Since complete inactivation of the *RB* gene appeared to be directly responsible for tumour formation, the *RB* gene was categorized as a tumour suppressor gene.

The *RB* gene, cloned in 1986, contains 27 exons dispersed over 200 kb of DNA encoding a mRNA of 4.7 kb (8, 10, 11). The *RB* gene was found to be expressed in all normal tissues, but absent or mutated in all RB tumours tested (11), which confirmed the hypothesis that inactivation of the *RB* gene is important in the pathogenesis of these tumours. The *RB* gene product (pRB) is a 110 kDa nuclear protein with DNA binding activity. It undergoes phosphorylation

in a cell-cycle dependent manner (12-14), with hypophosphorylated pRB present in G1 and hyperphosphorylated pRB from S to M.

Mutation or absence of pRB have been found not only in RB, but also in other tumours, including osteosarcoma, small cell lung carcinoma, breast, bladder and prostate cancer, and leukaemia (10, 15-21). These observations suggest a broad role for loss of functional pRB in the genesis of human tumours.

I. 2. The functions of RB protein (pRB)

I. 2. 1. The role of pRB in the cell cycle

Two lines of evidence suggest that pRB has an important role in the cell cycle. First, the phosphorylation of pRB changes during the cell cycle. During G1, pRB is predominantely hypophosphorylated, while in the S, G2 and M phases pRB exists in a progressively hyperphosphorylated form. In the study of primary T cells stimulated to enter the cell cycle, pRB has been shown to undergo phosphorylation in multiple stages during the cell cycle: it is first phosphorylated in mid G1; subsequent phosphorylation events occur in S and G2/M (22). Analysis of the phosphoprotein revealed that pRB is phosphorylated on serine/threomine residues by cyclin-dependent kinases (23). The second line of evidence comes from the discovery that microinjection of excess purified hypophosphorylated pRB into early G1 cells inhibits cell progression into S phase, but injection of similar amounts of hypophosphorylated pRB into cells arrested in late G1 or early S phase has no effect on DNA synthesis (24).

It is known that pRB can form complexes with several transforming proteins of DNA tumour viruses, such as SV40 large T antigen (25), adenovirus E1A (26) and human papilloma virus E7 (27). The SV40 large T antigen specifically binds to the hypophosphorylated form of pRB that exists in the G1

phase of the cell cycle, indicating that hypophosphorylated pRB is the active form (28, 29). Studies also showed that pRB-induced G1 arrest can be abolished by co-injection with T antigen (24). Because one of the functions of these viral proteins is to drive non-dividing cells into the cell cycle so as to permit viral replication, it is possible that binding viral oncoproteins to the active form of pRB inhibits pRB's ability to restrain cell division (27, 30, 31). Based on these findings, it was suggested that hypophosphorylated pRB possesses a suppression activity in cell growth that can be turned off in late G1 either by phosphorylation or by binding of viral oncoproteins.

The interaction of pRB and viral oncoproteins raised the possibility that pRB may function by binding to cellular proteins that are in turn responsible for cell cycle regulation at the G1/S boundary. Immunological and biochemical studies have provided evidence that pRB interacts with a number of cellular proteins implicated in the cell cycle (32-36). The best characterized cellular protein associated with pRB is the transcription factor E2F (32). E2F was originally identified as a cellular factor mediating the transcription of the adenovirus *E2* gene (33). In most cells, E2F is complexed with other cellular proteins, thus preventing E2F from forming a complex with the E4 protein (a 19 kDa product of the adenovirus *E4* gene activated during the early phase of viral infection) to activate the *E2* promoter. The adenovirus E1A protein can dissociate the binding between E2F and cellular proteins, allowing E2F to complex with E4 and bind to the *E2* promoter to induce transcription of the *E2* gene (33). E2F was also found to be required for transcription of cellular genes involved in growth control and DNA synthesis, such as *DHFR*, thymidine kinase, DNA polymerase α, cdc2, *MYCN* and *MYCC* (33, 36-39). These genes have been shown to have E2F binding sites in their promoter region. In addition.

expression of E2F has been found to induce quiescent cells to enter S phase (40). These observations suggest that E2F has a function in cell progression. By immunoprecipitation experiments, pRB has been found to be one of the cellular proteins associated with E1A (26), suggesting that pRB may be able to form a complex with E2F (32). Additional experiments demonstrated that pRB forms a complex with DRTF1, a protein related or identical to E2F (41). A "pocket" region has been identified at the carboxy terminus of pRB, which contains two domains (domain A from amino acids 394 to 571, domain B from amino acids 649 to 773). The pocket region is required for pRB interaction with E2F, E1A and SV40 large T antigen (42-44). When using this region of pRB as a probe to screen expression libraries, a number of clones that bind to pRB were obtained (45-48). One clone was human E2F-1 defined by its sequence-specific DNA binding and transcriptional activation properties (46-48). All these data suggest that E2F is a cellular target for pRB.

The interaction of E2F and pRB is regulated by the phosphorylation status of the pRB. E2F only binds to the hypophosphorylated form of pRB (32, 43). It has been suggested that in G1, pRB binds to E2F and prevents it from binding and activating transcription of other genes until the G1/S transition point, at which time pRB is phosphorylated by cyclin-dependent kinases and E2F is released. When E2F is free, it can activate transcription of genes required for S phase (32, 49). If pRB is absent or mutated, it can not bind to E2F and/or other similar factors. These factors can therefore continually activate transcription of the genes required for S phase and the cells do not arrest at the G1/S boundary. Through interaction with E2F, pRB acts as a regulator of the cell cycle by overseeing the G1/S transition, regulating a large number of genes required for cell cycle transition and also playing a role in the differentiation of some cell types (14, 50, 51).

I. 2. 2. The role of pRB in transcription

Since pRB is a nuclear protein and has DNA-binding activity *in vitro*, pRB has also been proposed to function as a transcriptional factor regulating expression of certain genes. Transfection experiments have provided evidence supporting this possibility.

pRB has been reported to negatively regulate the transcription of the *c-fos* gene (52). Expression of *c-fos* was shown to be essential for the entrance of quiescent cells into the cell cycle and its promoter appeared to be an attractive target for suppressors of cell growth (53, 54). Human or mouse *c-fos* promoters linked to a reporter gene (CAT) were transfected into NIH/3T3 cells alone or in combination with an expression vector carrying RB cDNA. The results indicated that the level of *fos* expression in cotransfected cells was reduced compared to the level in the cells transfected with the CAT reporter constructs alone (52). In an effort to map the regulatory elements within the *c-fos* promoter that are responsible for pRB repression, 5' deletion mutants of the *c-fos* promoter were tested for their activity in the presence of RB cDNA. A 31 nucleotide region between -102 and -71 in the human *c-fos* promoter was defined as being essential for transcriptional repression by pRB. This region has been termed retinoblastoma control element (RCE) (52). Since *c-fos* is one component of the heterodimeric transcription factor AP-1 believed to be involved in regulating a set of early response genes required for cell growth, it is proposed that pRB may also indirectly regulate the expression of other components of AP-1, such as *c-jun*, *junB* and *junD* by downregulating AP-1 activity through repression of *c-fos* transcription (52).

pRB has also been shown to positively regulate the transcription of TGF-β *1* and TGF-β2. These two genes belong to the family of transforming growth factor-βs (TGF-βs) that have a function in the regulation of cell division, cell

differentiation and are the most potent growth-inhibitory polypeptides known for a wide variety of cell types (55-57). Sequence homology to the *c-fos* RCE has been found in the TGF-β*1* gene promoter and transfection analysis has demonstrated that pRB can regulate TGF-β*1* gene expression positively or negatively via this site depending on the cell types, suggesting that the regulation of TGF-β*1* expression by pRB is cell-type specific (58). For TGF-β2, the pRB induced transcription activation is through interaction with the transcription factor ATF-2 (59).

Another gene known to be regulated by pRB is *MYCC*. *MYCC* is a member of the *MYC* oncogene family. It encodes a nuclear phosphoprotein that functions as a transcription regulator involved in the control of cell proliferation and differentiation (60). Studies showed that pRB can downregulate transcription of *MYCC*. This repression is believed to be mediated by TGF-β*1* (61). A TGF-β control element (TCE) has been found in the *MYCC* promoter by deletion analysis (62). The sequence of TCE is homologous to RCE. Both of them are GC rich sequences. Co-transfection of RB cDNA and *MYCC* promoter also showed that the TCE is necessary for downregulation of *MYCC* by pRB (62).

I. 2. 3. Are other genes involved in RB tumour formation?

All the results from the studies described above suggest that inactivation of pRB may directly or indirectly affect numerous genes and their products. In order to understand the RB tumorigenic state at the molecular level, it is important to study not only the primary events involved in the transforming process, but also the effect of pRB inactivation on the expression of other genes in the cells. One approach is to identify genes that are selectively or preferentially expressed in RB cells. For this purpose, a RB-enriched cDNA library was prepared from RB522A and Y79 by hybridizing RB522A and Y79

cDNAs to excess poly $(A)^+$ RNA from fetal tissues and a human duodenum carcinoma cell line (HuTu). The double-strand hybrid was separated from the single-strand cDNA by hydroxyapatite chromatography. The nonhybridized single-strand cDNA was selected for construction of the library (63). This procedure resulted in the depletion of transcripts that are commonly present in normal and tumorigenic proliferating cells. A differential strategy was then used to screen the cDNA library by hybridizing replica filters with labelled cDNA derived from either RB522A/Y79 or fetal tissue/HuTu poly $(A)^+$ RNA, respectively. Clones hybridizing preferentially to the RB522A/Y79 cDNA were selected, purified and used for further analysis. Using this strategy, a transcript that is preferentially expressed in a subset of RB cell lines has been identified. A search for nucleotide sequence homology using the GenBank DNA database revealed that this transcript encodes a member of the DEAD box protein family (63). This DEAD box protein has been named DDX1(DEAD box 1).

II. The DEAD box protein family

The DEAD box protein family is a rapidly growing family of proteins, which contains more than 30 members at the present time. DEAD box proteins have been identified in a wide range of organisms from bacteria to humans (64, 65). This family was originally identified based on their amino acid sequence homology to the eukaryotic translation initiation factor eIF-4A. This family is characterized by eight highly conserved amino acid motifs (Figure 1), one of which is the DEAD motif (DEAD standing for Asp-Glu-Ala-Asp in their single letter amino acid code) that gives the family its name (65). Members of this family have been shown to have RNA unwinding activity.

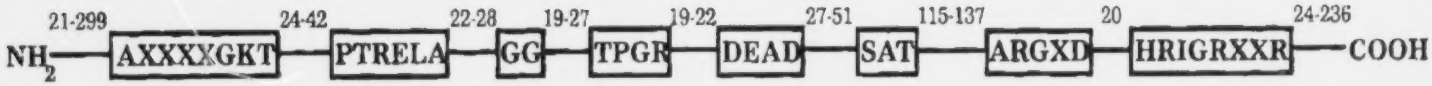


Figure 1. Schematic representation of highly conserved amino acid motifs in the DEAD box protein family (64). The eight conserved motifs are in bold and boxed. The numbers between the boxes indicate the distance in amino acid residues.

II. 1. Conserved motifs in the DEAD box protein family

All members of the DEAD box protein family have been shown to contain eight highly conserved motifs in their amino acid sequences. Of these eight motifs, only four have been shown to have biochemical activities. They are motifs one (AXXXXGKT), five (DEAD), six (SAT) and eight (HRIGRXXR). Using site-directed mutagenesis, Sonenberg and his colleagues have carried out in vitro functional studies of these four highly conserved regions in mouse eIF-4A (66). Furthermore, Linder and Schmid have performed an in vivo mutational analysis of motifs one, five and six in yeast Tif1 and Tif2 (Tif1 and Tif2 have been shown to be functional homologues of mouse eIF-4A) (67). The data from their studies have demonstrated that motifs one, five, six and eight are involved in the ATP binding, ATP hydrolysis and RNA unwinding reactions. The first motif (AXXXXGKT) has been described as an ATPase A motif and shown to be involved in ATP binding (66, 68, 69). The alanine (A) in this motif is conserved in most of the members of the family (64). Substitution of this alanine residue with a valine or an aspartic acid is either lethal to the cell or results in a slower growth rate, whereas replacement by glycine has no effect on cell growth (67). UV cross-linking assays showed that changing alanine to valine results in a dramatic decrease in ATP cross-linking while changing lysine (K) to asparagine results in almost complete abrogation of ATP binding. The fifth motif (DEAD) represents an ATPase B motif (68). Evidence from mutational analysis indicates that this motif is involved in ATP hydrolysis (66, 70) and some studies suggest that it may also be involved in ATP binding (65, 66, 68, 70). The D-E residues in this motif are highly conserved not only in the DEAD box family, but also in many proteins involved in DNA and RNA replication (71, 72). Mutations in this region abolish ATP hydrolysis, indicating that they are required for ATP hydrolysis (66). The sixth motif (SAT) has been shown to be essential for RNA unwinding and a

mutant with residues AAA instead of SAT was incapable of unwinding duplex RNA, but still capable of binding RNA in an ATP-dependent manner (66). The eighth motif (HRIGRXXR) located in the C-terminus is believed to be involved in ATP hydrolysis and has been identified as a critical region for RNA binding. It has been shown *in vitro* that any mutations generated by site-directed mutagenesis in this region of eIF-4A drastically reduces RNA cross-linking as well as the rate of ATP hydrolysis activity (66, 73). RNA helicase activity of eIF-4A was also affected by the mutations in this last conserved motif (66, 73).

All these studies show that these four highly conserved motifs (motifs one, five, six and eight) are important for the function of the DEAD box proteins as defined by ATPase and RNA helicase activities. The functions of the remaining motifs that are conserved in the DEAD box family have not been determined yet.

Although all members of the DEAD box family contain eight conserved motifs, individual members have C-terminal and N-terminal regions that have little in common. These regions have been suggested to be associated with the specific function of individual proteins (64, 65, 66).

II. 2. The functions of the DEAD box proteins

Two well studied DEAD box proteins, eIF-4A and p68 (a mammalian nuclear protein), have been shown to have RNA-dependent ATPase and ATP-dependent RNA helicase activities *in vitro* (73-77). eIF-4A is known to complex with eIF-4F (which is composed of the cap-binding protein eIF-4E and a 220 KDa protein) and bind to the 5' end of mRNA. In the presence of another translation initiation factor eIF-4B and using the energy generated from ATP hydrolysis, this complex will unwind secondary structure in the 5' untranslated region of eukaryotic mRNA to allow ribosome binding and scanning for the AUG codon (78-80). An *in vitro* RNA-unwinding assay has revealed that eIF-4A unwinds

RNA in both the 3'- to 5' and 5'- to 3' directions (73). Based on these results, it is often assumed that all members of the DEAD box protein family have RNA-dependent ATPase and ATP-dependent RNA helicase activities and that these proteins have a function in manipulation of RNA secondary structure (63, 64). Studies have demonstrated that DEAD box proteins are implicated in various cellular functions. Some are involved in general cellular processes, such as translation initiation [eIF-4A in mouse (73, 81), Tif1/Tif2 in yeast (66, 83)], RNA splicing [PRP5, PRP28 and SPP81 in yeast (84-86)] and ribosomal assembly [SrmB in *E.coli* (87)]. Other members of this family are known to be tissue- or stage- specific. Their cellular functions are not clear, but some are believed to play a regulatory role in cell growth and development. For example, murine PL10 mRNA is expressed only in the male germ line and its product has been suggested to have a specific role in translational regulation during spermatogenesis (88). The *Drosophila* genes, *vasa* and *ME31B*, are maternally expressed and their products are believed to be involved in embryogenesis (89-92). p68 was identified because of its specific immunological cross-reaction with the SV40 large T antigen (93). It is expressed in all dividing eukaryotic cells, but not in quiescent cells (93). The nuclear location of p68 undergoes dramatic changes during the cell cycle. It stays in the nucleoplasm during interphase and enters the nucleoli during telophase (78). Although p68 has been shown to have RNA-dependent ATPase and ATP-dependent RNA helicase activities *in vitro* (76, 77), little is known concerning its cellular functions. It has been suggested that p68 may be required for the formation of nucleoli and may also have a function in the regulation of cell growth and division (76-78).

In addition to roles in translation initiation and mRNA splicing, DEAD box proteins are also involved in stabilizing mRNA (94). In *E.coli* cells, replacing the host RNA polymerase with the faster bacteriphage T7 enzyme results in a ~100

fold decrease in the production of β-galactosidase. But when vectors carrying *E.coli* DEAD box protein genes, either *SrmB* or *deaD* were introduced into *E.coli* cells containing bacteriphage T7 enzyme, there was overproduction of β-galactosidase and cell growth was not affected. It is proposed that overexpression of DEAD box proteins stabilized the β-galactosidase transcripts by protecting them from endonucleases and resulted in efficient translation. (94).

II. 3. Regulation of DEAD box proteins

The mechanisms responsible for regulation of the DEAD box proteins are not yet understood. However, a recent study has shown that p68 contains a region of sequence homology to the conserved protein kinase C (PKC) phosphorylation site and calmodulin (CaM) binding domain of neural-specific proteins (95). The CaM binding domain is also known as the IQ domain (96). It has been shown that CaM binding to the IQ domain of neuromodulin decreases the rate of phosphorylation by PKC, and phosphorylation conversely prevents CaM binding to the IQ domain (97). The IQ domain also has been implicated as a region for interaction with other Ca^{++}-binding proteins (98, 99). A study has shown that after removal of bound RNA from p68 using ribonuclease, p68 can be phosphorylated by PKC in a Ca^{++}-dependent manner *in vitro* and phosphorylation results in reduced ATPase activity, suggesting that PKC phosphorylation blocks RNA stimulation of ATPase activity. In addition, the ATPase activity of p68 can be significantly inhibited by CaM binding and this inhibition is both Ca^{++}- and CaM dose-dependent (95). Since the levels of nuclear CaM in the cells is normally very high and Ca^{++} or CaM alone has no effect on ATPase activity of p68, it was suggested that the ATPase activity of p68 is likely inhibited by PKC and/or by Ca^{++}/CaM through the IQ domain (95). Because ATPase activity is required for p68 unwinding activity, it is proposed

that the RNA helicase activity of p68 may be regulated by PKC and/or Ca^{++}/CaM during the cell cycle (95).

III. DDX1

DDX1 is a human DEAD box protein identified in RB cells by differential screening of a RB-enriched cDNA library. All eight conserved amino acid motifs characteristic of DEAD box proteins are present in the deduced amino acid sequence of DDX1 (63). Studies from this lab have shown that the *DDX1* gene is expressed at high levels in two of six RB cell lines tested -- RB522A and Y79, as well as in IMR32 (a neuroblastoma cell line), but not in a variety of tumorigenic cell lines from different origins, such as malignant glioma, bladder and colon cancers (63). Tissue distribution of DDX1 mRNA indicates that it is expressed at higher levels in tissues of neuroectodemal origin, including retina, brain and spinal cord, than in other tissues (63). Southern blot analysis of the *DDX1* gene revealed that there is gene amplification in the cell lines that have elevated levels of DDX1 mRNA, indicating that gene amplification may be one of the factors responsible for the higher expression of the *DDX1* gene in these cell lines (63). Interestingly, the *MYCN* gene is also found to be amplified in these same cell lines (RB522A, Y79 and IMR32) (63) and recent studies have shown that both *DDX1* and *MYCN* genes are amplified in some primary neuroblastoma tumours and cell lines (100-103), suggesting co-amplification of these two genes in these cell lines and tumours (63, 100-103). Chromosome mapping studies revealed that *DDX1* and *MYCN* colocalize to chromosome band 2p24. Furthermore, HSR (homogeneously staining region) analysis has indicated that amplified copies of *DDX1* and *MYCN* are located on the same HSR region on chromosome 1 in the neuroblastoma cell line IMR32 (63, 104, 105). These data strongly support the possibility of *DDX1* and *MYCN* co-amplification.

The biochemical and cellular functions of DDX1 protein remain unknown, but some hypotheses have been generated based on what is known concerning DDX1 and other members of the DEAD box protein family. Based on sequence homology with other members of the DEAD box protein family, we hypothesize that DDX1 protein has RNA-dependent ATPase and ATP-dependent RNA helicase activities. In addition, since there is a similar level of homology between DDX1 and all other members of DEAD box protein family, DDX1 may be involved in either RNA splicing, initiation of translation or other activities yet to be defined (63). Furthermore, computer search for amino acid sequence homology of the region between the first two conserved motifs in DDX1 has shown a high level of homology to a subregion of hnRNP U (heterogeneous nuclear ribonucloprotein U) (106). HnRNP U is a member of the hnRNP family. This family has been proposed to have a function in the processing of hnRNA to mRNA (107). The hnRNP U has RNA binding activity, and binds to hnRNA *in vivo* and to both RNA and ssDNA *in vitro* (108). The presence of a hnRNP U homologous region in DDX1 indicates that DDX1 may have a role in RNA processing (106). Finally, based on the fact that *DDX1* and *MYCN* genes are co-amplified in some RB and neuroblastoma cell lines, it has been postulated that there may be interaction between the products of *DDX1* and *MYCN* genes in these cell lines. The DDX1 protein may play a role in MYCN mRNA processing. Alternatively, MYCN protein, a transcriptional regulator (109, 110), may play a role in the regulation of DDX1 RNA (100).

IV. The aim of this study

The overall objective of this study was to characterize the *DDX1* gene in RB and neuroblastoma cell lines. There were two aspects to this study: the first

was to clone the 5' end of the DDX1 cDNA and identify the transcriptional start site, and the second was to characterize the regulatory elements involved in the regulation of *DDX1* transcription.

The results from this study may contribute to a better understanding of the mechanism of transcriptional regulation of the *DDX1* gene and may also help to further investigate the possible relationship between *DDX1* and *MYCN* genes at the molecular level.

CHAPTER II.

MATERIALS AND METHODS

Library Screening

Positive phage were purified using three rounds of screening. In the primary screening, the library was plated at a density of 5 X 10^4 p.f.u. (plaque forming units) per 150 mm plate. *E. coli* LE392 bacterial host cells were used to screen genomic libraries. *E. coli* BB4 was used to screen cDNA libraries. A total of 1 x 10^6 phage particles were analyzed in the genomic library screening and 3 x 10^5 phage particles in the cDNA library screening. Host cells (600 μl) were incubated with phage at 37°C for 20 min (this allows the phage to infect the host cells) and plated on 150 mm plates. The plates were incubated at 37°C for 5 to 6 hours. The plates were then chilled at 4°C overnight. Phage were transferred to a nitrocellulose filter by placing a dry filter (130 mm, Millipore HA) on the surface of the plate for 3 min. After asymmetrically marking each filter and the corresponding plate for orientation using waterproof ink, the filter was lifted from the plate with forceps, denatured in 1.5 M NaCl and 0.5 M NaOH, neutralized in 1.5 M NaCl and 0.5 M Tris HCl (pH 7.5), and rinsed in 2X SSC and 0.2 M Tris HCl (pH 7.5) for 5 min, respectively. All filters were air dried for 1/2 hour and then baked at 80°C in a vacuum oven for 2½ hours.

The filters were pre-hybridized at 42°C for 4 to 5 hours in 50% formamide, 5X SSC, 5X Denhardt's, 50 mM NaH_2PO4 (pH 6.5) and 250 μg/ml sonicated salmon sperm DNA. The pre-hybridization buffer was then removed and 10 ml of hybridization solution containing 50% formamide, 5X SSC, 1X Denhart's solution, 20 mM NaH_2PO_4, 10% dextran sulfate, denatured ^{32}P-labelled probe and 100 μg/ml salmon sperm DNA was added. Hybridization was carried out at 42°C overnight. Following hybridization, the filters were washed three times in 2X

SSC and 0.1% SDS, and two times in 0.1X SSC and 0.1% SDS. The washed filters were air dried and exposed to X-ray film at -80°C with intensifying screens. Positive clones were picked with toothpicks and transferred to a grided 85 mm plate containing a lawn of host cells. At least 50 plaques were picked from each positive clone area in order to increase the chance of getting positive clones. The plaques underwent secondary screening as detailed for primary screening except that 80 mm nitrocellulose filters (Millpore HA) were used to transfer plaques.

The tertiary screening involved removing an agar plug containing the positive plaque, transferring the agar plug to a sterile eppendorf tube containing 500 μl of SM buffer (0.1 M NaCl, 0.01 M $MgSO_4$, 0.05 M Tris HCl pH 7.5 and 0.01% gelatin) and 35 μl of chloroform, and releasing the phage particles from the agar plug. This phage stock was stored at 4°C. Diluted phage stock was incubated with 200 μl of host cells at 37°C for 20 min and plated on a 85 mm plate. Using toothpicks, individual plaques were picked and transferred to a grided 85 mm plate containing a lawn of host cells. Phages were transferred to nitrocellulose filter and underwent tertiary screening.

Positive plaques from the tertiary screening were stored as individual phage stocks in 500 μl of SM buffer and 35 μl of chloroform at 4°C.

For the cDNA library screening, pBluescript phagemids containing DNA inserts of interest were prepared from phage stocks by *in vivo* excision according to the protocol supplied by Stratagene. For the genomic library screening, bacteriophage containing DNA inserts of interest were prepared from phage stocks as described (111) and then digested with restriction enzyme. The digestion products were separated on a 1% agarose gel and Southern blotted in order to select DNA fragments of interest. The DNA fragments were then purified and subcloned into pBluescript.

Preparation of DNA Inserts

DNA fragments were generated by restriction enzyme digestions, separated by electrophoresis on 1% agarose or 5% polyacrylamide gels, and isolated from the gels by electroelution in 1X TBE at 160 volts for 2 to 3 hours. The DNA was extracted with phenol and chloroform, and then precipitated with 1/10 volume of 5 M NaCl and 2 volumes of ethanol at -20°C for at least 1 hour. The DNA pellets were resuspended in sterile H_2O and stored at -20°C.

Labelling of DNA Fragments

The DNA fragments used as probes in library screenings or Southern blotting were labelled using nick translation as described (113). The DNA fragments used as probes in the EMSA experiments were labelled as described (114). Single-end labelling of DNA fragments used as probes in DNase I footprinting was performed as described (115).

DNA Sequencing

DNA sequencing was by the dideoxynucleotide chain-termination method with T7 DNA polymerase (Pharmacia) as modified for double-stranded DNA templates (116, 117). For sequencing of large DNA fragments, a sequential deletion strategy was used to obtain overlapping sequences (118).

Promoter Sequence Homology Search

Searches for promoter sequence homology were performed using the eukaryotic promoter database at the National Center for Biotechnology Information through the BLAST network service. In addition, Dr. Rufus Day (Cross Cancer Institute) performed a search for transcription factor binding sites using his specially designed program.

Secondary Structure Prediction

The secondary structure prediction for the sequence at the 5' end of DDX1 cDNA was performed using the 'foldrna' program from the GCG (Genetics

Computer Group) software manual version 7.3.1. 1993 (119). This program is based on the method described in references 120 and 121.

RACE (Rapid Amplification of cDNA Ends)

A 5'-AmpliFINDER RACE kit was purchased from CLONTECH. Using poly $(A)^+$ RNA as template and nested gene-specific primers, the RACE experiments were carried out according to the protocol supplied by the manufacturer (Figure 2). Briefly, using 2 μg of poly $(A)^+$ RNA from RB522A as template, 1st-strand cDNA was synthesized with a *DDX1* gene-specific primer. The RNA template was then hydrolyzed with NaOH and excess primer was removed using a glass matrix support. Purified cDNA was precipitated with ethanol. A single-stranded AmpliFINDER anchor oligonucleotide that contains an EcoRI site was ligated to the 3'-end of the cDNA using T4 RNA ligase. PCR amplification was then performed using ligated cDNA as template and the two primers, AmpliFINDER anchor primer and a nested *DDX1* gene-specific primer. RACE PCR products were analysed by electrophoresis on a 1% agarose gel and Southern blotting. PCR products of interest were purified, restriction enzyme digested and then cloned into pBluescript. DNA sequence analysis of both strands of the DNA insert was as described in DNA sequencing.

Southern Blot Analysis

RACE PCR products were run on a 1% agarose gel in TAE buffer (40 mM Tris-acetate, 1 mM EDTA pH 7.2). The DNA gel was then denatured and neutralized in 0.5 N NaOH/1.5 M NaCl and 3 M NaCl/0.5 M Tris-HCl (pH 7.5), respectively. Transfer of DNA to nitrocellulose and conditions for prehybridization and hybridization were as described (122).

Poly $(A)^+$ RNA Preparation

The RB522A cell line was obtained from Dr. Brenda Gallie (Hospital for Sick Children, Toronto). The IMR32 cell line was obtained from the American

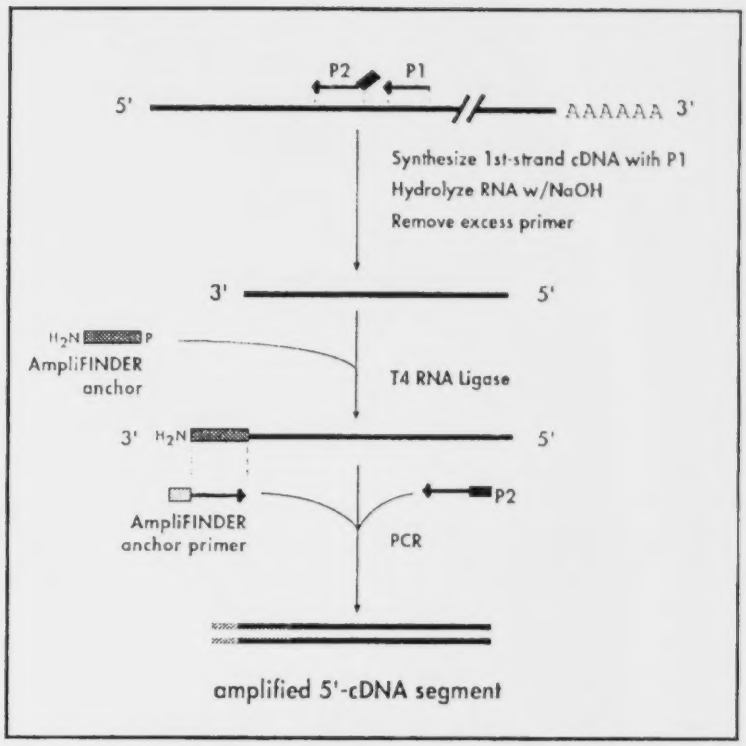


Figure 2. Schematic representation of amplification of the 5'-end of a gene using the 5'-AmpliFINDER RACE method (supplied by CLONTECH). P1 and P2 are nested gene-specific primers. After the 1st-strand cDNA is synthesized using poly(A)+ RNA as template and P1 as primer, the AmpliFINDER anchor is ligated to the 3'-end of 1st-strand cDNA using T4 RNA ligase. AmpliFINDER anchor primer, which contains a *EcoR*I site, and P2 are then used for PCR amplification. The black square at the 3'-end of P2 indicates a restriction enzyme site that will be used in cloning the PCR product.

Type Culture Collection. The RB(E)-2 tumour was obtained from the Royal Alexandra Hospital here in Edmonton. Cells were cultured in DMEM containing 10% FCS (fetal calf serum), 10 μg/ml insulin and 5 X 10^{-5} M β-mercaptoethanol. Poly $(A)^+$ RNA was extracted from these cells using the hot phenol method and oligo(dT)-cellulose chromatography as described (123).

Primer Extension

A primer extension system from Promega was used in this study. The experiments were carried out according to the protocol supplied by the manufacturer. Briefly, poly $(A)^+$ RNA was isolated from RB522A, IMR32 and RB(E)-2 cells, respectively. An oligonucleotide complementary to the sequence at the 5' end of the DDX1 cDNA was labelled with T4 polynucleotide kinase (BioLabs) and [γ-^{32}P] ATP (DuPont, 3000 Ci/mmol, 10 mCi/ml). The labelled primer was annealed to 2 μg of each poly $(A)^+$ RNAs at 45°C for 90 min. The extension reaction followed at 42°C for 60 min using AMV reverse transcriptase (Promega). The primer extension products were heated at 90°C for 10 min and then separated on a 8% denaturing polyacrylamide gel. A G+A sequencing ladder served as the size standard. The gel was dried at 80°C in an vacuum dryer for 50 min and then exposed to X-ray film at -80°C.

DNase I Hypersensitivity Assay

Nuclei were isolated from RB522A cells as described (124). DNase I was added to 100 μl (~100 μg) aliquots of nuclei to a final concentration of 0.1-25 μg/ml. The reactions were then incubated at 37°C for 10 min. The negative controls in which nuclei were not treated with DNase I, were incubated at 4°C and 37°C for 10 min, respectively. Digestion was stopped by adding 200 μl of buffer containing 10 mM Tris-HCl (pH 8.0), 2.5 mM EDTA, 1% SDS and 50 μg/ml proteinase K and incubated at 37°C overnight. Genomic DNAs were then extracted by chloroform and phenol and precipitated with ethanol. The purified

DNA (10 μg per sample) was digested with appropriate restriction enzymes. The DNA fragments were separated by electrophoresis on a 1% agarose gel and *Hind*III digested λ DNA was used as a marker. Southern blotting was carried out using radioactively labelled probes prepared by nick translation.

Preparation of Nuclear Extracts

Nuclear extracts were prepared from RB522A and IMR32 cells according to the method described by Guérin et al (125). Briefly, about 5 X 10^8 frozen cells from RB522A or IMR32 were thawed on ice. 4 ml of NE1 buffer [10 mM Hepes (pH 7.9), 1.5 mM $MgCl_2$, 10 mM KCl, 0.2 mM EDTA and 25% glycerol] was added and the cells were homogenized using a 15 ml dounce. Nonidet P-40 was then added to a final concentration of 0.2%. Following another round of homogenization, nuclei and cell debris were centrifugated at 1300 rpm for 10 min at 4°C using a HB-4 rotor. The supernatant was discarded. Nuclei were washed with 5 ml of NE1 buffer and centrifugated as above. After the supernatant was discarded, the nuclei were gently resuspended with 1 "Packed Cell Volume" of NE1 buffer and then 4 M KCl was slowly added to a final concentration of 350 mM. The nuclei were incubated on ice for 10 min, then transferred to a 7 ml dounce and homogenized ~10 strokes. The homogenate was transferred to ultracentrifuge tubes and spun at 30,000 rpm for 3 hours at 4°C using a TST 60.4 rotor. The supernatant was dialyzed twice, 30 min each time at 4°C against 500 ml of DNase I buffer (50 mM KCl, 4 mM $MgCl_2$, 20 mM KPO_4, 1 mM β-mercaptoethanol and 20% glycerol), after which the supernatant was transferred to eppendorf tubes and centrifugated at full speed for 5 min in a microcentrifuge at 4°C. The supernatant was quick-frozen in 50-100 μl aliquots in liquid nitrogen and stored at -80°C. A small aliquot of the supernatant was used to determine the protein concentration according to the Bradford assay (126).

EMSA (Electrophoretic Mobility Shift Assay)

The DNA fragments used as probes in EMSA were labelled with [α-^{32}P] dCTP (DuPont, 800 Ci/mmol, 10 mCi/ml) using Klenow polymerase for filling in at restriction enzyme sites. EMSA was performed essentially as described by Garabedian et al (127). Briefly, 1-2.5 μg of nuclear extract were incubated with 1-2 μg of poly (dI-dC) in a final volume of 20 μl containing binding buffer (20 mM Tris-HCl pH 7.9, 50 mM NaCl, 2 mM $MgCl_2$, 1 mM EDTA, 10% glycerol, 1 mM dithiothreitol) for 20 min at room temperature. The labelled probe was added and incubated for another 10 min at room temperature. Competition binding experiments were performed by first incubating the competitor with the nuclear protein extract and poly (dI-dC) in binding buffer for 20 min at room temperature. The labelled probe was then added and the sample was incubated for an additional 10 min. The reaction mixture was loaded onto a 6% non-denaturing polyacrylamide gel and resolved by electrophoresis at 160V at room temperature for 2 to 3 hours. The gel was dried for ~1 hour at 80°C and then exposed to X-ray film.

***In vitro* DNase I Footprinting Assay**

In vitro DNase I footprinting experiments were performed essentially as described (127). Briefly, DNA fragments used as probes in the DNase I footprinting experiments were end-labelled on either the coding and/or the noncoding strand with [α-^{32}P] dCTP (Amersham, 3000 Ci/mmol, 10 mCi/ml) by a fill-in reaction using Klenow polymerase. A second restriction enzyme was used to obtain the appropriate end-labelled DNA fragments. The digested products were separated on a 5% polyacrylamide gel. The end-labelled fragments were cut out from the gel and purified. DNase I digestions were carried out in a 50 μl volume containing 25 mM Hepes (pH7.6), 23 mM KCl, 6 mM $MgCl_2$, 0.48 mM EDTA, 10% glycerol, 0.5 mM dithiothreitol, 2 μg of poly (dI-dC) and end-labelled

probe (10 fmole, 10,000-30,000 cpm) with 30 μg of nuclear extracts. The mixtures were incubated on ice for 15 min and then at room temperature for 2-3 min, after which 50 μl of start buffer (10 mM $MgCl_2$, 5 mM $CaCl_2$) was added. Fresh DNase I diluted in sterile H_2O was added and digestion was allowed to proceed for 60 sec at room temperature. The amount of DNase I was adjusted to produce an even pattern of partial cleavage products. Reactions were stopped by adding 100 μl of stop solution (200 mM NaCl, 20 mM EDTA, 10% SDS). Samples were then extracted with phenol-chloroform, and the DNAs were precipitated with 2 volumes of ethanol. The DNA pellets were dried and resuspended in sequencing loading buffer (95% formamide, 10 mM NaOH, 1 mM EDTA, 0.025% bromophenol, 0.025% xylene cyanol), and the samples were denatured at 90°C for 3 min and then loaded onto a 7% polyacrylamide-8M urea sequencing gel. The gel was dried and exposed to X-ray film at -80°C. Areas protected from DNase I digestion and their sequences were determined by parallel running of a G+A sequencing reaction and by comparison with a negative control reaction in which no nuclear extract was added.

CHAPTER III.

CLONING THE 5' END OF DDX1 cDNA AND IDENTIFICATION OF THE TRANSCRIPTION INITIATION SITE

III. 1. Introduction

Obtaining a complete cDNA sequence is critical for gene regulation and expression studies. By differential screening of a RB-enriched cDNA library, a ~2.4 kb sequence of DDX1 cDNA (this sequence is referred to as the DDX1 2.4 kb cDNA in this thesis) was generated from a number of overlapping cDNA clones. The presence of a poly (A) tail at the 3' end of the DDX1 2.4 kb cDNA indicated that it was complete at its 3' end. The predicted amino acid sequence of DDX1 revealed an open reading frame from position 1 to 2217 of the DDX1 2.4 kb cDNA (106). However, based on the predicted size of the DDX1 mRNA determined by Northern blotting, it was estimated that a maximum of 400 bp of 5' DDX1 cDNA sequence was lacking. Extensive screening of two RB cDNA libraries failed to generate clones containing the 5' end of DDX1 cDNA.

In order to obtain the full length sequence of the DDX1 cDNA, two strategies were used. First, Northern analysis showed that DDX1 mRNA is preferentially expressed in fetal brain, suggesting that screening of a fetal brain cDNA library might represent a way to obtain clones whose sequence extended to the 5' end of DDX1 cDNA. A human fetal brain cDNA library was therefore purchased for this purpose. Second, we used the RACE (Rapid Amplification of cDNA Ends) strategy to attempt to obtain the 5' end of DDX1 cDNA. This method is designed to specifically amplify the 5' ends of cDNAs and is believed to be the preferred way to overcome the difficulties of cloning the 5' ends of cDNAs caused by secondary structure at the 5' end of mRNA. Once the complete

sequence of the DDX1 cDNA is obtained, the primer extension strategy can be used to verify the exact transcription start site.

III. 2. Results

Cloning the 5' end of DDX1 cDNA. A human fetal brain cDNA library (week 18 of gestation) was purchased from Stratagene. Approximate 500,000 plaques were screened using a ^{32}P-labelled probe (a ~320 bp DNA fragment from the 5' end of the DDX1 2.4 kb cDNA). Clones hybridizing to the probe were selected, purified and the pBluescript phagemids containing DNA inserts were rescued by *in vivo* excision. Upon sequencing of purified DNA from these clones, we were able to extend the DDX1 cDNA sequence to the 5' end by ~100 bp (this sequence will be referred to as DDX1 2.5 kb cDNA).

In the RACE experiment, we designed two nested *DDX1*-specific primers, P1 and P2 (Figure 3). These primers were complementary to the sequence at the 5' end of DDX1 2.4 kb cDNA. Using 2 μg of poly (A)$^+$ RNA from RB522A as template, 1st-strand cDNA was synthesized with P1. In order to obtain efficient cDNA synthesis, the poly (A)$^+$ RNA was first incubated at 65°C for 15 min to reduce secondary structure at the 5' end of DDX1 mRNA. For the same reason, the reverse transcriptase step in the cDNA synthesis was performed at a relatively high temperature (52°C). The RNA template was then hydrolyzed using NaOH and cDNA was synthesized. A single-stranded AmpliFINDER anchor oligonucleotide containing an *EcoR*I site was ligated to the 3' end of the cDNA using T4 RNA ligase. PCR amplification was carried out using ligated cDNA as template and the two primers: AmpliFINDER anchor primer and P2. In control reactions, either AmpliFINDER anchor primer and P1 or only AmpliFINDER anchor primer were used in the PCR reactions. An additional control reaction was carried out in the absence of ligated cDNA and using only AmpliFINDER

5'CTCTGTAGCTGTGACCCTGATACCGCGTGGTGTGCTCCGAACACATGGT
P4
GCCCAGAACGAAGGCGGCGTCCAGAAGCCCTAGGTCCCAGAGGTCGCTCAG
P3
CGGCAGGCGCATAAGGCGGGGCCGGCGCGGGCCTTTCCTTCCATCGGAACC

GTTCTCCCGGGGCTGAGTCCCTGCCCGGACTCCGAACGCCGAAGACCAGGG

GCCGGAAGCGCGCGCCGCCACTGCCACGCCGTGTCAGTCGGGAGGGAGGGA

GCGAGCAGGCGAAGCCGCGGAGGACGGGGTGAAGATGGCGGCCTTCTCCGA

GATGGGTGTAATGCCTGAGATTGCACAAGCTGTGGAAGAGATGGATTGGCT

CCTCCCAACTGATATCCAGGCTGAATCTATCCCATTGATCTTAGGAGGAGG
P2
TGATGTACTTATGGCTGCAGAA...3'
P1

P1, P2... gene-specific primers used in the first RACE experiment. The primers used are complementary to the underlined sequences.
P3, P4... gene-specific primers used in the second RACE experiment. The primers used are complementay to the underlined sequences.

Figure 3. Sequence of the 5' end of the DDX1 2.7 kb cDNA. The sequence shown in bold represents the additional sequence obtained using the RACE strategy. This sequence extended ~230 bp upstream of DDX1 2.5 kb cDNA.

anchor primer in the PCR reaction. PCR products were electrophoresed on a 1% agarose gel and Southern blotted. The filter was hybridized using a ^{32}P-labelled probe (a ~100 bp DNA fragment from the 5' end of DDX1 2.5 kb cDNA). A PCR product of ~400 bp was found to hybridize to the probe (Figure 4). The DNA from this band was purified, digested with *EcoR*I/*EcoR*V and cloned into pBluescript digested with the same enzymes. Several clones were sequenced from both strands. The longest clone contained an insert of ~370 bp that extended ~230 bp upstream of DDX1 2.5 kb cDNA (Figure 3) (this sequence is referred to as DDX1 2.7 kb cDNA). Based on this sequence, we designed an additional two 5' end-specific primers, P3 and P4 (Figure 3), and repeated the RACE experiment. P3 was used in cDNA synthesis and P4 was used in PCR amplification. The sequence obtained from the second RACE experiment did not extend further upstream. Therefore, both RACE experiments revealed the same sequence at the 5' end of the DDX1 cDNA.

When the 'foldrna' computer program was used to analyse 303 nucleotides at the 5' end of the DDX1 2.7 kb cDNA, it revealed an extensive secondary structure (Figure 5). Based on these data, we speculated that our original difficulties in cloning the 5' end of the DDX1 cDNA may have been due to the secondary structure at the 5' end of DDX1 mRNA that prevented the progression of the reverse transcriptase through this region during first strand cDNA synthesis.

Identification of the transcription initiation site of DDX1. The transcription initiation site of DDX1 was determined by primer extension. Poly $(A)^+$ RNA was prepared from RB522A, IMR32 and RB(E)-2 cells, respectively. RB522A and IMR32 cell lines were chosen for their overexpression of DDX1 mRNA, while RB(E)-2 was chosen because it expresses DDX1 mRNA at normal levels. The primer used in the primer extension was a 21 oligonucleotide that

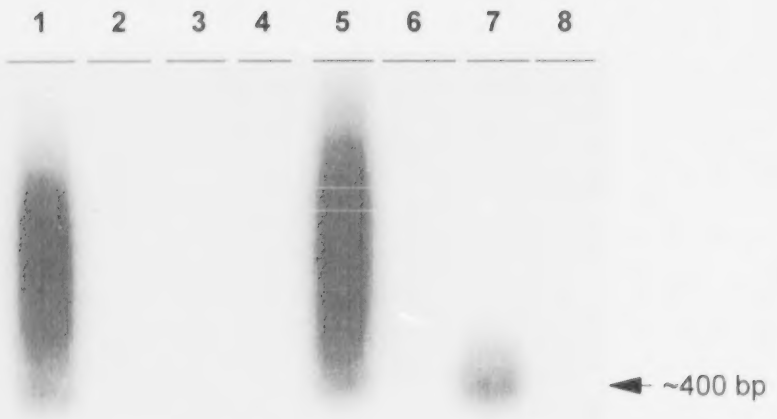


Figure 4. Southern blot analysis of PCR products from the RACE experiments. 1st-strand cDNA was synthesized using poly(A)+ RNA from RB522A as template and gene specific primer P1 (Figure 4). Anchor primer (Figure 2) and gene specific primer P2 (Figure 4) were used for PCR amplification. In control reactions, either anchor primer and gene specific primer P1 were used or only the anchor primer was used in the PCR reactions. 1 μl of each PCR product was electrophoresed on a 1% agarose gel and a ~100 bp radioactively labelled cDNA fragment derived from the 5' end of DDX1 2.5 kb cDNA was used as the probe in Southern blots. Lanes 2-4, 6 and 8 were loaded with 1μl of each control samples [lane 2, using anchor primer and primer P1; lane 3, using only anchor primer; lane 4, using only anchor primer and without template DNA; lanes 6 and 8, using anchor primer and primer P1, but the ligated cDNA used in the PCR reactions was diluted (1/10 in lane 6 and 1/100 in lane 8)]. Lanes 1, 5 and 7 were loaded with 1 μl of RACE PCR product using anchor primer and P2, respectively [(the ligated cDNA used in the PCR reactions were undiluted (lane 1), 1/10 diluted (lane 5) and 1/100 diluted (lane 7)]. The arrow indicates the location of the ~400 bp PCR product that hybridized to the probe.

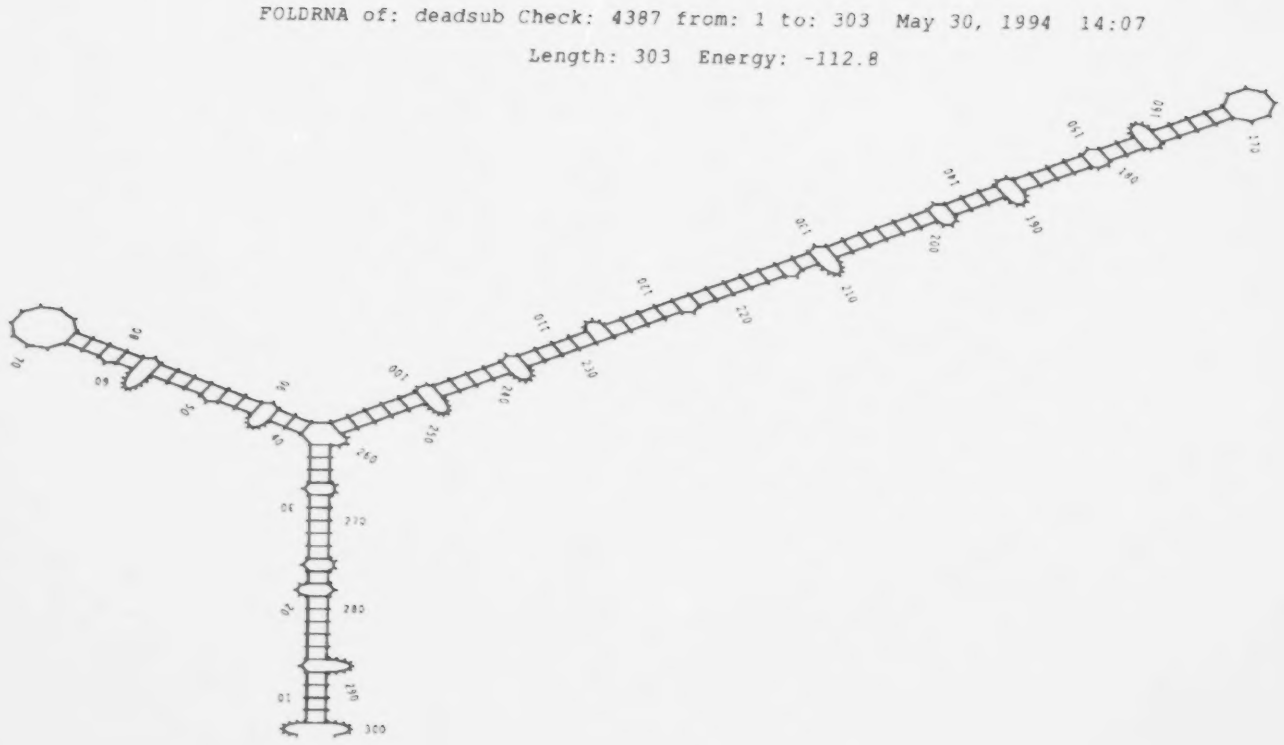


Figure 5. The structural diagram generated by foldrna program represents the most stable secondary structure at the 5' end of the DDX1 2.7 kb cDNA.

was complementary to a sequence near the 5' end of the DDX1 2.7 kb cDNA (Figure 6), and labelled with T4 polynucleotide kinase and [γ-^{32}P] ATP. The labelled primer (~200 fmol, 150,000 to 250,000 cpm) was annealed to 2 μg of poly $(A)^+$ RNA from RB522A, IMR32 and RB(E)-2 cells, respectively, and incubated at 45°C for 90 min. The cDNA was extended 5' of the primer with reverse transcriptase at 42°C for 60 min. The products of the primer extension were separated on a 8% sequencing gel using a G+A sequencing ladder as size marker. The result of a typical primer extension is shown in Figure 7. No extension products were detected in RB(E)-2. In the RB522A cell line, three major extension products were observed at 46, 43 and 40 bp upstream from the primer. The first two corresponded to positions 6 and 3 nucleotides upstream of the 5' end of the DDX1 2.7 kb cDNA and the third one corresponded exactly to the 5' end of the DDX1 2.7 kb cDNA (Figure 6). One minor band located at 45 bp upstream from the primer was also detected in the RB522A cell line (Figure 7). This band corresponded to a site 5 nucleotides upstream of the 5' end of the DDX1 2.7 kb cDNA (Figure 6). In the IMR32 cell line, we did not observe any extension products corresponding to those seen in the RB522A lane, despite increasing the exposure time and repeating primer extension experiments under different conditions. But there is a evidence (data not shown) that extension products identical to those observed in the RB522A cell line are found in another neuroblastoma cell line, BE(2)-C, also shown to overexpress DDX1 mRNA (100). Based on these data, we believe that there are three major transcription initiation sites in *DDX1* (Figure 6). We have designated the nucleotide at which the first transcription initiation site is located as "+1" (Figure 6).

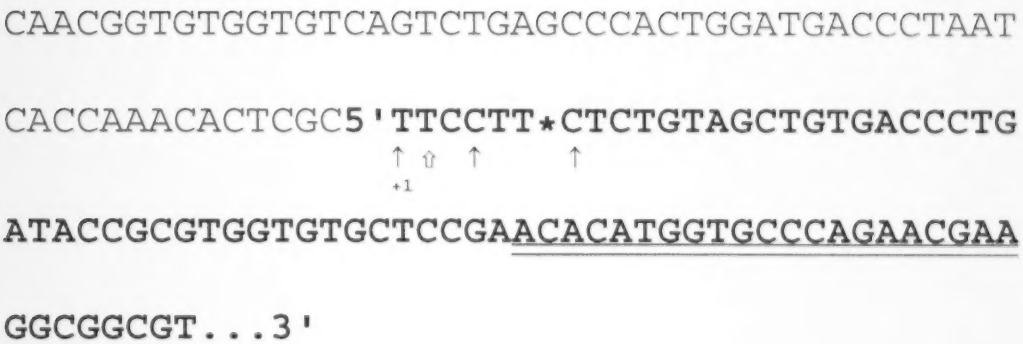


Figure 6. Transcription start sites of *DDX1* as determined by primer extension. Three major transcription initiation sites are indicated by "↑" and one minor site is indicated by "⇧". The nucleotide indicated by the first "↑" is designated as "+1", which is located 6 nucleotides upstream of the 5' end of the DDX1 cDNA obtained using the RACE strategy (the latter is denoted by "*"). The primer used in the primer extension is complementary to the sequence that is double underlined.

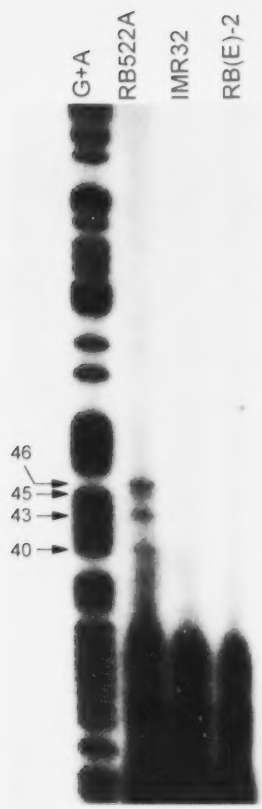


Figure 7. Mapping of *DDX1* transcription initiation sites by primer extension. A radioactively labelled primer (Figure 6) was annealed to 2 μg of poly (A)+ RNA from RB522A, IMR32 and RB(E)-2 cells, respectively, and extended to the 5' end using reverse transcriptase. The products were run on a 8% denaturing polyacrylamide gel using a G+A sequencing ladder as size marker.The extension products detected are indicated by arrows and their positions upstream from the primer used in primer extension are numbered in nucleotides, which are determined by comparison to the G+A sequence ladder.

CHAPTER IV.

CHARACTERIZATION OF THE REGULATORY ELEMENTS OF THE DDX1 GENE

IV. 1. Introduction

Gene expression is controlled by regulation of gene transcription that involves specific interactions between transcription factors and regulatory elements. In this study, characterization of regulatory elements that participate in the transcriptional regulation of the *DDX1* gene will help us to understand how the expression of this gene is regulated. Since the regulatory elements are usually located upstream of the transcription initiation site, we initiated this study by screening a human genomic library to identify and clone the 5' flanking sequence of *DDX1*.

We have used different techniques to identify the regulatory elements of *DDX1*, including DNase I hypersensitivity, EMSA (Electrophoretic Mobility Shift Assay) and DNase I footprinting. DNase I hypersensitivity experiments have proven to be a valuable tool for detection of alterations in chromatin structure that are associated with the regulatory regions of transcribed genes (124, 128-130). These regions are "hypersensitive" to DNase I digestion and therefore called DNase I hypersensitive (HS) sites. HS sites are believed to occur as the consequence of transcription factors binding to specific DNA sequences (130) and have been mapped to promoters (regulatory elements that are proximal to the transcription initiation site) and enhancers (more distally located elements that usually function in a position and orientation independent manner) (131). Mapping of DNase I HS sites can provide general information regarding the location of regulatory elements in a large segment of DNA flanking the gene of interest.

EMSA and DNase I footprinting are two methods commonly used to detect specific interactions between transcription factors and *cis*-regulatory elements (132-135). Transcription factors are sequence-specific DNA binding proteins, which are involved in the regulation of gene transcription by binding to sequence-specific DNA located at promoters or enhancers, resulting in either activation or repression of gene transcription.

EMSA provides a simple and sensitive method for identifying nuclear proteins that bind to specific sequences of a gene (127). It relies on the ability of a protein to bind to a radioactively labelled DNA fragment *in vitro*, followed by electrophoretic separation of DNA-protein complexes from unbound DNA on a non-denaturing polyacrylamide gel (133, 134). DNA-complexes migrate more slowly and give retarded bands on the gel. Using this method, protein-DNA interactions in a given DNA sequence region can be determined. In DNase I footprinting, the binding of transcription factors to an end-labelled DNA fragment protects the DNA domain that interacts with the transcription factor from digestion by DNase I, resulting in protected regions represented as absent bands on denaturing polyacrylamide gels (127). By using a G+A sequence ladder in DNase I footprinting, the sequences recognized by the transcription factors can be accurately identified.

Another method that can be used to identify regulatory elements is DNA transfection assay using CAT reporter gene. This is a functional assay to biologically assess the regulatory activity of DNA flanking the gene of interest. DNA fragments from the 5' flanking region, which contain putative regulatory elements, are cloned into a pCAT vector at a position upstream of the *CAT* gene that encodes the enzyme chloramphenicol acetyl transferase (CAT). Eukaryotic cells are then transfected with these constructs. As the pCAT vector lacks eukaryotic promoter and enhancer sequences, CAT activity will be driven by the

regulatory elements located in the inserted genomic DNA. CAT activity in cultured cells is measured by the conversion of [^{14}C] chloramphenicol to its 1-acetyl and 3-acetyl derivatives, which can be analyzed either by using thin layer chromatography (TLC) or by liquid scintillation counting (LSC) of CAT reaction products (136). Because *CAT* is a bacterial gene, eukaryotic cells contain no or very low background of CAT activity. The CAT activity detected in this assay therefore reflects the activity of regulatory elements located in the inserted DNA.

IV. 2. Results

Genomic cloning. A human genomic library was purchased from CLONTECH. Approximate 1 X 10^6 plaques were screened using a ^{32}P-labelled probe consisting of a ~170 bp DNA fragment derived from the 5' end of the DDX1 2.7 kb cDNA. Two clones that hybridized to the probe were selected and purified by three rounds of screening. Genomic DNA from these two clones were digested with *EcoR*I, *Pvu*II and *Bgl*II, respectively. The products of these digestions were separated on a 1% agarose gel and Southern blotted. The filter was hybridized using the same ~170 bp probe as the one used in genomic library screening. Two DNA fragments derived from one clone, an *EcoR*I fragment of ~9 kb and a *Bgl*II fragment of ~4 kb, hybridized to the probe. These two fragments were purified and subcloned into pBluescript. In order to obtain the complete sequence of the ~4 kb *Bgl*II fragment, a sequential deletion strategy was used. The pBluescript containing this insert was double-digested with *Kpn*I and *Xho*I. *Kpn*I digestion created a 3'-single-stranded overhang that prevented the vector sequence from being digested by exonuclease III, while *Xho*I digestion produced a 5' overhang from which exonuclease III initiated digestion to create deletions in the insert. By controlling the time of exonuclease III digestion, sequential deletion of the insert was obtained. After digestion with

mung bean nuclease, the DNAs were re-ligated. Overlapping sequences were obtained by sequencing plasmid DNAs containing various size inserts. Sequencing of the ~4 kb *Bgl*II fragment indicated that it contains ~2.1 kb of 5' flanking sequence. The sequence immediately upstream of the transcription initiation site did not reveal a typical TATA box sequence. However, a TATA box sequence (TATAAA) was found at position -260 to -255 bp. Eukaryotic promoter homology search for the ~2.1 kb of 5' flanking sequence was performed using the eukaryotic promoter database at the National Center for Biotechnology Information through the BLAST network service. Two areas with homology to human Alu sequences were identified at -1883 to -1576 bp and -1049 to -603 bp, respectively. The region from -315 to -272 bp (Figure 8) was homologous to the *MYCN* promoter region from -214 to -170 bp [66% (30/45) nucleotide identity] (137, 138).

Mapping DNase I hypersensitive (HS) sites in *DDX1*. Three DNase I hypersensitivity experiments were performed to map the DNase I HS sites in the 5' flanking region, as well as in the first exon and first intron of *DDX1*. Nuclei were prepared from RB522A cells and treated with increasing concentrations of DNase I. The genomic DNA was then purified and digested with three different enzymes. The genomic DNA fragment produced from the restriction enzyme digestion, along with subfragments produced by DNase I digestion, were detected using specific probes mapping 5' or 3' of the fragment. Through analysis of the size of the bands detected, possible DNase I HS sites were mapped. In the first DNase I HS experiment, the DNA was digested with *Bgl*II, which generates a ~4.2 kb fragment from -2.1 kb to +2.1 kb. This fragment, along with the partial digestion products produced by DNase I digestion, were detected with probe A (as indicated in Figure 9). Four DNase I HS sites were identified in this region, HS sites I and II were mapped within the first intron, HS

5'....GACACTCTGGGGCAGGCTACAGATAGCACAGCACGCGCCCTCGCATTGCGCAGA

TGTGGCTCAGATCCCACGCTGCCATTCGGTAGATAATGTGTCTTTAGATAAGTGACTTAA

CTTTCCCGAATTTTAGAGTCCTCATCTAAAGATTCTAATACCTTCCTGACAGAGTGAGTG

TCTGGTTGGGAGGATTCCTCATTTACTTTCC-315CCATCTCTCTGCAAAGAGAAAGGTGTTAA

AATTGGCATGGAAGTC-272AAACCCGAAATTATAAAAGCGACATCAAACATGGGTGAAGGGAG

GGAGACCTAAATCATTACTTAGTTTAAAAAAAAAATGTAACAAATACCTACTGGCTGCCA

GGCACTCACTGGCATTGGTGCTCAGGGCAATGAATCCAGTCCATCTCTCCGGGTCATTCA

CACACCAAGGCCCTGCTCACCTTTGGAAAGCTGGGGGACTGAGTGGGGAGGGCACGGTGT

GGTGTCAGCCTGAGCCCACTGGATGACCCTAATCACCAAACATCTGC+1**TTCCTTCTCTGTA**

GCTGTGACCCTGATACCGCGTGGTGTGCTCCGAACACATGGTGCCCAGAACGAAGGCGGC

GTCCAGAAGCCCTAGGTCCCAGAGGTCGCTCAGCGGCAGGCGCATAAGGCGGGGCCGCGC

GGGCCTTTCCTTCCATCGGAACCGTTCTCCCGGGGCTGAGTCCCTGCCCGGACTCCGAAC

GCCGAAGACCAGGGGCCGGAAGCGCGCGCCGCCACTGCCACGCCGTGTCAGTCGGGAGGG

AGGGAGCGAGCAGGCGAAGCCGCGGAGGACGGGGTGAAGATGGCGGCCTTCTCCGGTGCG

TTTGTGGAAACTCTGGGGGTGGTGGGGCGGCTTGTTGCACGCCTCAGCCCTGGGGCGCCC

GCGGAGGCTAAAGGGGCGGGAGCGGACAGGGGTCAGGGAGACAGAAGGGCGCTGTCCGCT

GCAGTCCCTGA....3'

Figure 8. Sequence showing the *MYCN* promoter homology region in *DDX1*. The sequence homology to the *MYCN* promoter is double underlined and extends from -315 to -272 bp. The sequence shown in bold represents the first exon of *DDX1* and the transcription initiation site is at "+1".

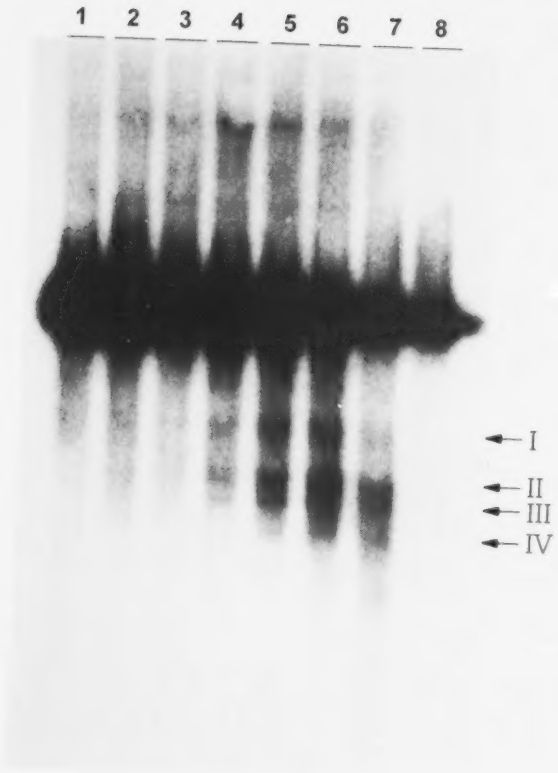
site III was within the first exon and HS site IV was mapped at a region immediately upstream of the transcription initiation site (Figure 9). The second DNase I HS experiment was carried out using *Hind*III, which generates a ~5.2 kb fragment from -1.1 kb to +4.1 kb. Probe B (as indicated in Figure 10) was used to detect the 5.2 kb fragment and subfragments resulting from DNase I digestion. Three DNase I HS sites were identified. Two of them (II and IV) had previously been identified using probe A. One new site (V) was mapped within the first intron (Figure 10). To examine the DNase I HS sites in the far upstream region of *DDX1*, *Sma*I was used in the third DNase I HS experiment. *Sma*I digestion generated a fragment of ~25 kb spanning from -25 kb to +160 bp of *DDX1*. This fragment and other subfragments generated by DNase I were detected by two 3' probes, probe C and probe D (as indicated in Figure 11). In the Southern analysis, instead of one ~25 kb fragment expected to be detected from *Sma*I digestion, there were two fragments at ~25 kb, which is probably caused by polymorphism in the *Sma*I site in the genome. Two DNase I HS sites (VI, VII) were identified in this region which mapped to -3.4 kb and -16.5 kb, respectively (Figure 11).

The DNase I HS sites identified by DNase I HS experiments are summarized in Figure 12.

EMSA and DNase I footprinting assays. To investigate the interaction of transcription factors and sequence specific DNA, we performed EMSA and DNase I footprinting experiments. Four DNA fragments (Figure 13) covering the region from -577 bp to +431 bp of *DDX1* were isolated by restriction enzyme digestions and nuclear extracts were prepared from RB522A and IMR32 cells.

In EMSAs, each of four DNA fragments, which are fragment 1 (-577/-203), fragment 2 (-202/-36), fragment 3 (-35/+163) and fragment 4 (+164/+431) (Figure 13) was radioactively labelled and incubated with nuclear extracts from either

Figure 9. Mapping of DNase I hypersensitive (HS) sites in a region from -2.1 kb to +2.1 kb of *DDX1*. Nuclei prepared from RB522A were untreated (negative control) or treated with DNase I and DNAs were purified. 10 μg of each DNA was digested with *Bgl*II, generating a ~4.2 kb fragment from -2.1 kb to +2.1 kb shown in the diagram . Lanes 1 and 2 represent DNA from nuclei that were not digested with DNase I but were incubated at 4°C and 37°C for 10 min, respectively. Lanes 3-8 represent DNA from nuclei treated with DNase I at 0.1, 1, 2.5, 5, 10 and 25 μ g/ml at 37°C for 10 min, respectively. Probe A, a ~420 bp *Dra*I/*Hind*III fragment shown in the diagram, was used to detect the ~4.2 kb fragment and subbands resulting from DNase I digestion. The subbands, corresponding to DNase I HS sites, are marked by roman numerals and their possible locations in *DDX1* are indicated by arrows in the diagram. The black box represents the first exon of *DDX1*. The transcription initiation site is at "+1". Restriction enzyme digestion sites shown are B (*Bgl*II), D (*Dra*I) and H (*Hind*III).


1
2
3
4
5
6
7
8
I
II
III
IV

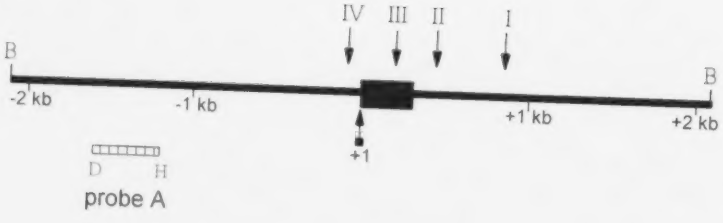
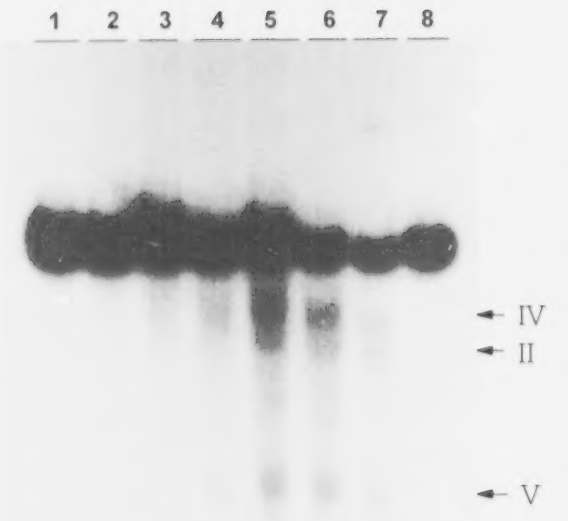
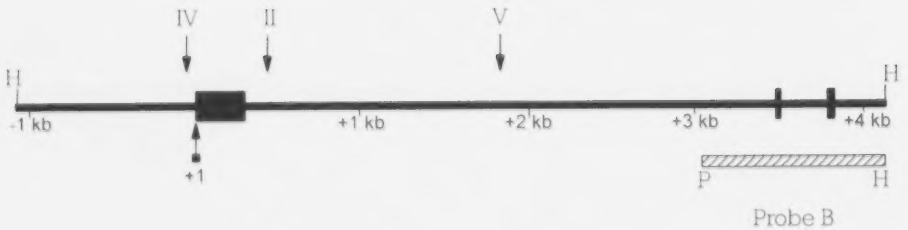

IV
III
II
I
B
B
-2 kb
-1 kb
+1 kb
+2 kb
+1
D
H
probe A

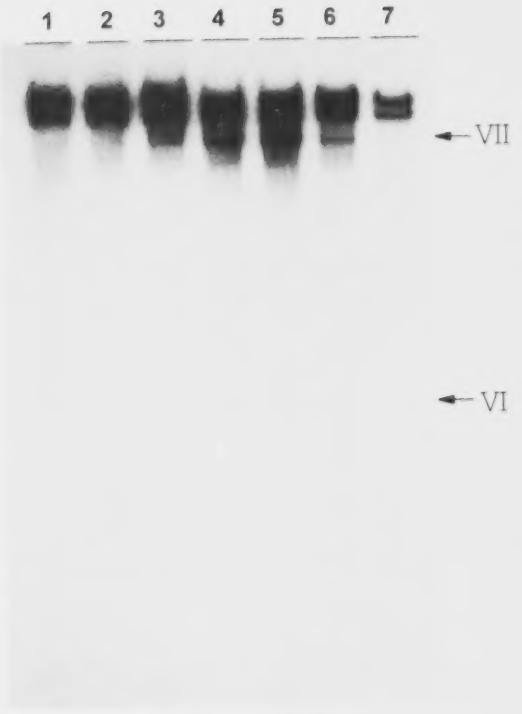
Figure 10. Mapping of DNase I HS sites in a region from -1.1 kb to +4.1 kb of *DDX1*. Nuclei of RB522A cells were treated with varying concentrations of DNase I. After DNA purification, 10 μg of each sample was digested with *Hind*III, which generated a fragment of ~5.2 kb from -1.1 kb to +4.1 kb. Probe B, a ~1.1 kb *Hind*III/*Pvu*II fragment shown in the diagram, was used to detect the ~5.2 kb fragment and partial digestion products generated by DNase I. DNA samples loaded in each lane are as described in the legend to Figure 9. The subbands representing DNase I HS sites are marked by roman numerals. Sites II and IV were previously identified using probe A. The possible locations of DNase I HS sites in *DDX1* are indicated by arrows in the diagram. The black boxes represent the first three exons of *DDX1* and the transcription initiation site is at "+1". Restriction enzyme sites shown are H (*Hind*III) and P (*Pvu*II).

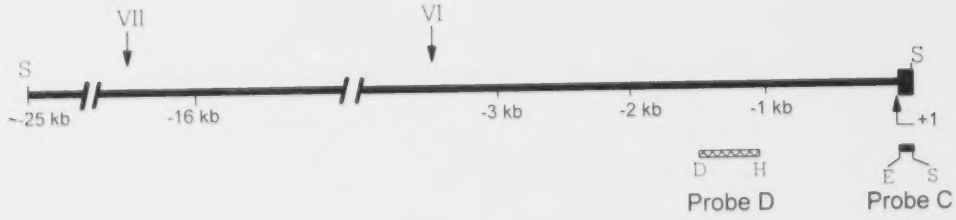




Probe B

Figure 11. Mapping of DNase HS sites in the upstream flanking region of *DDX1*. 10 μg of DNA from DNase I-untreated and -treated nuclei of RB522A were digested with *Sma*I, which generates a ~25 kb DNA fragment from -25 kb to +160 bp of *DDX1*. Lane 1 was loaded with DNA from untreated nuclei (at 37°C for 10 min). Lanes 2-7 were loaded with DNA from nuclei treated at 0.1, 1, 2.5, 5, 10 and 25 μg/ml DNase I at 37°C for 10 min, respectively. Probe C (a ~160 bp *EcoR*I/*Sma*I) and probe D (a ~420 bp *Dra*I/*Hind*III) shown in the diagram were used to detect the DNA fragments resulting from *Sma*I digestion and subbands generated by DNase I. Two bands at ~25 kb were observed instead of the one band expected in this region. The presence of two bands likely represents a polymorphism at *Sma*I site. The subbands are marked by roman numerals and their possible locations in *DDX1* are indicated by arrows in the diagram. The black box represents part of the first exon of *DDX1*. Restriction enzyme digestion sites shown are S (*Sma*I), D (*Dra*I), H (*Hind*III) and E (*EcoR*I).


1
2
3
4
5
6
7
VII
VI


VII
VI
S
S
~-25 kb
-16 kb
-3 kb
-2 kb
-1 kb
+1
D
H
E
S
Probe D
Probe C

Mapping of DNase I hypersensitive (HS) sites in *DDX1*

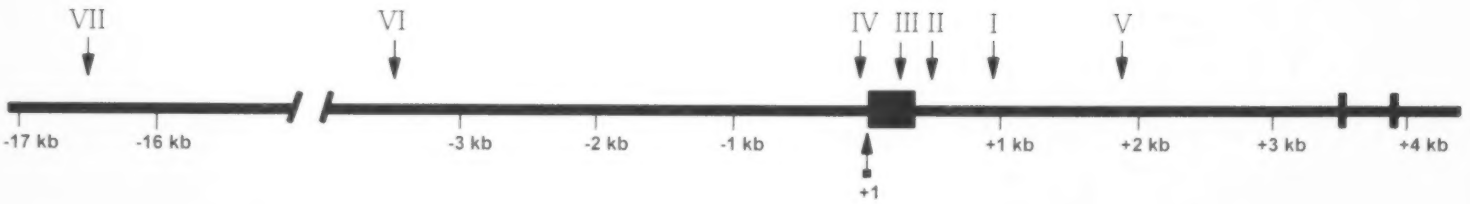


Figure 12. Summary of DNase I HS sites obtained from the DNase I HS experiments. The possible location of DNase I HS sites in *DDX1* are indicated by arrows and numbered by roman numerals. The black boxes represent the first three exons of *DDX1*. The transcription initiation site is located at "+1".

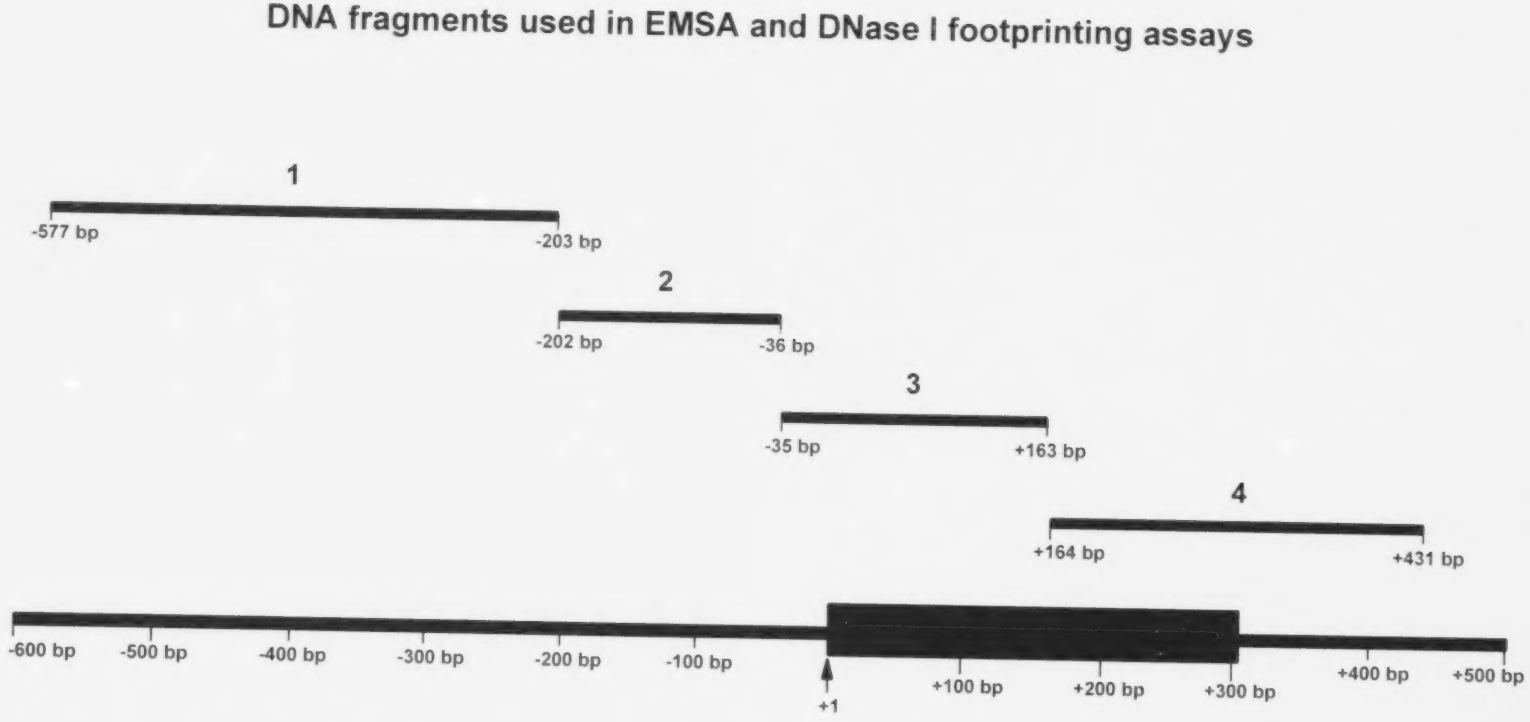
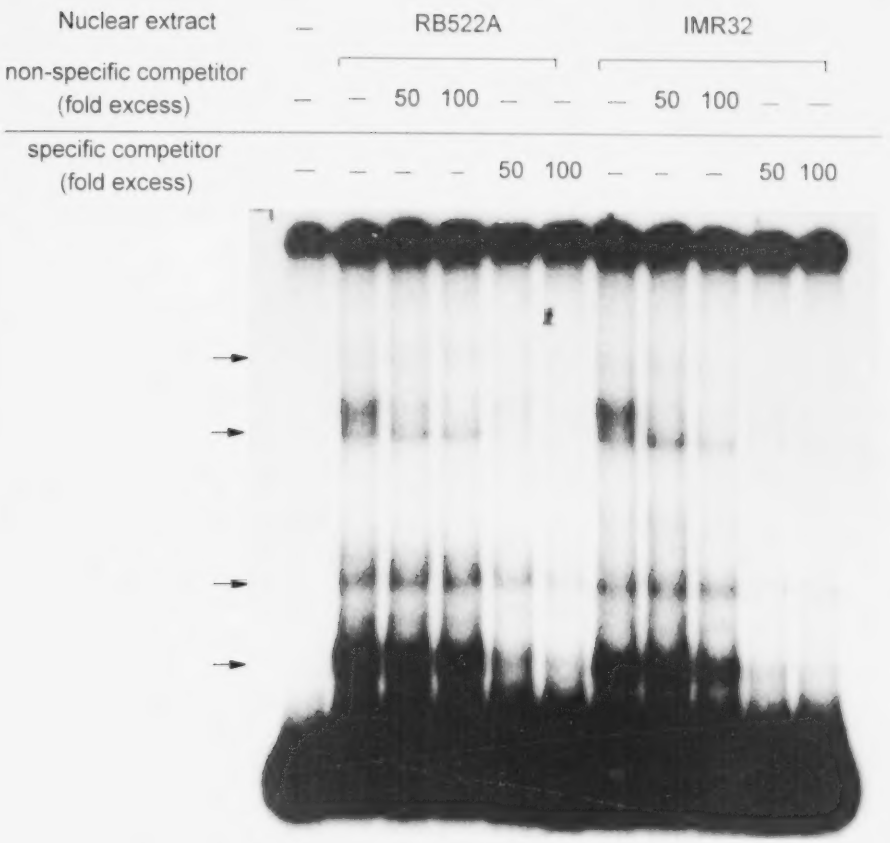


Figure 13. Schematic representation of DNA fragments used in EMSA (DNA fragments 1-4) and DNase I footprinting (DNA fragments 1-3). Four DNA fragments from *DDX1* were isolated by restriction enzyme digestions.These are numbered 1, 2, 3 and 4 in the diagram. They represent the region from -577 bp to +431 bp of *DDX1*. The first exon of *DDX1* is represented by the black box and the transcription initiation site is indicated by the arrow.

RB522A or IMR32. In negative control reactions, no nuclear extract was added. The DNA-protein complexes were separated from free labelled DNA and detected as retarded bands on nondenaturing polyacrylamide gel. The specificity of these complexes was demonstrated by competition reactions using nonspecific competitor (a ~300 bp fragment from a cDNA clone isolated from a day 3½ chick retina library) and specific competitor (same DNA fragment as the one used as the probe but unlabelled). As shown in Figure 14, five complexes were detected when using DNA fragment 1 (-577/-203) as the probe in EMSA. These complexes were present using both RB522A and IMR32 nuclear extracts. Competition reactions indicated that four of these complexes were produced by sequence-specific DNA-binding proteins which bind specifically to the sequences in DNA fragment 1. One complex was non-specific. We reached these conclusions based on the following observations: (i) when 100-fold excess of unlabelled DNA fragment 1 were added to the reaction, the retarded bands resulted from specific DNA-protein interactions either disappeared or the density of the bands was reduced significantly; (ii) when non-specific competitor was added at 100-fold excess compared to the labelled DNA fragment 1, there was no effect on these bands; (iii) for a non-specific DNA-protein complex, the retarded band was removed not only when specific competitor was added but also when non-specific competitor was added.

Using DNA fragment 2 (-202/-36) as the probe in EMSA experiments, three retarded bands were observed (Figure 15). All were present using both RB522A and IMR32 nuclear extracts. Competition reactions showed that only one of these three bands resulted from specific DNA-protein interaction. Using DNA fragment 3 (-35/+164) as the probe in EMSA experiments revealed five DNA-protein complexes (Figure 16). Competition reactions indicated that four of them were specific. The fifth complex was considered non-specific because it

Figure 14. Analysis of the interaction between nuclear proteins and DNA fragment 1 (-577/-203) of *DDX1* by EMSA. 1 μg of nuclear extract from either RB522A or IMR32 cells was incubated with 3 ng of labelled DNA fragment 1. Unlabelled DNA fragment 1 was used as the specific competitor and a ~300 bp cDNA fragment from a day 3½ chick retina cDNA library was used as the non-specific competitor. The specific and non-specific competitors were added in 50 and 100 fold excess compared to the labelled DNA fragment 1. Four specific retarded bands, which resulted from nuclear proteins binding to sequence specific DNA, are indicated by arrows. The diagram shows the location of DNA fragment 1 (-577 bp to -203 bp) in *DDX1*. The black box represents the first exon of *DDX1* and the transcription initiation site is at "+1".


Nuclear extract
–
RB522A
IMR32
non-specific competitor
(fold excess)
– – 50 100 – – – 50 100 – –
specific competitor
(fold excess)
– – – – 50 100 – – – 50 100

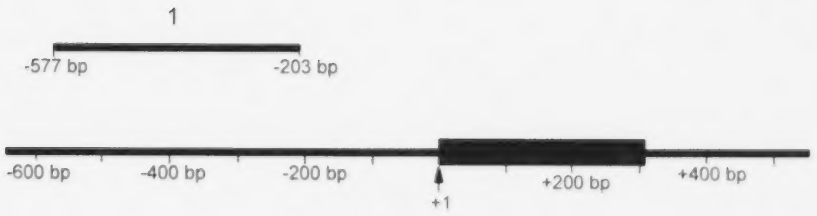
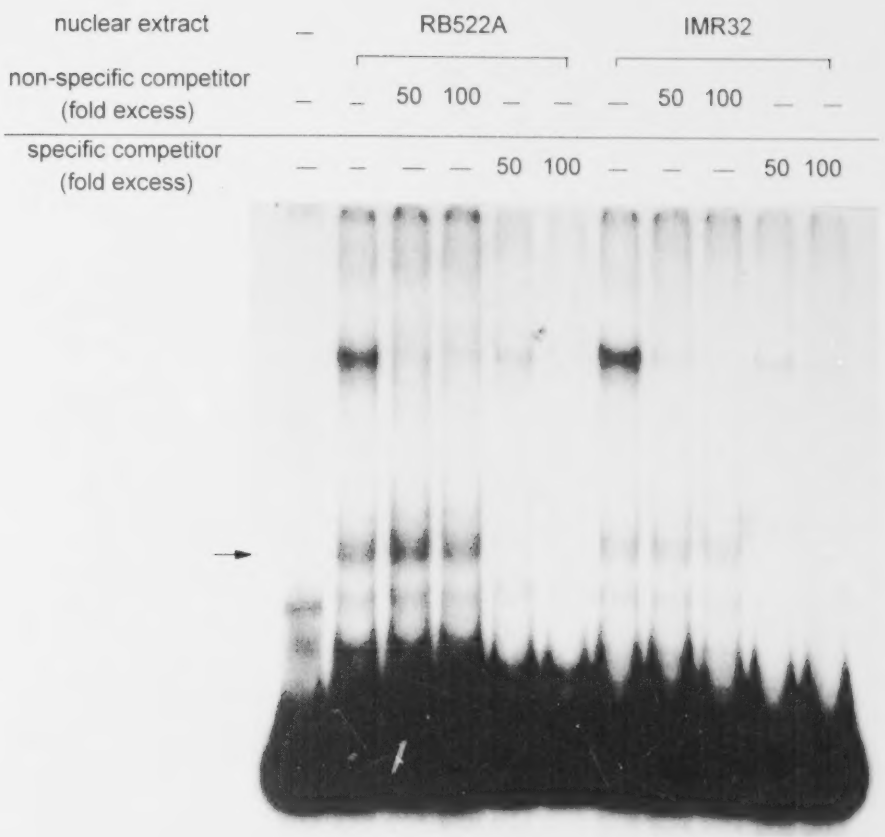
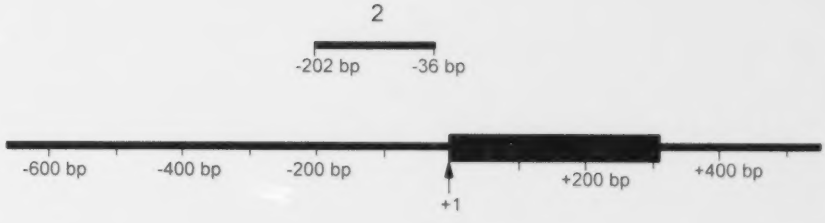

1
-577 bp
-203 bp
-600 bp
-400 bp
-200 bp
+1
+200 bp
+400 bp

Figure 15. Analysis of the interaction between nuclear proteins and DNA fragment 2 (-202/-36) of *DDX1* by EMSA. 1.5 μg of nuclear extracts from either RB522A or IMR32 cells were incubated with 4 ng of labelled DNA fragment 2. Unlabelled DNA fragment 2 was used as specific competitor. The DNA fragment used as non-specific competitor is as described in the legend to Figure 14. Specific and non-specific competitors were added in 50 and 100 fold excess compared to the labelled DNA fragment 2. One specific retarded band resulting from nuclear protein binding to sequence specific DNA is indicated by an arrow. The diagram shows the location of DNA fragment 2 (-202/-36) in *DDX1*. The black box represents the first exon of *DDX1* and the transcription initiation site is at "+1".

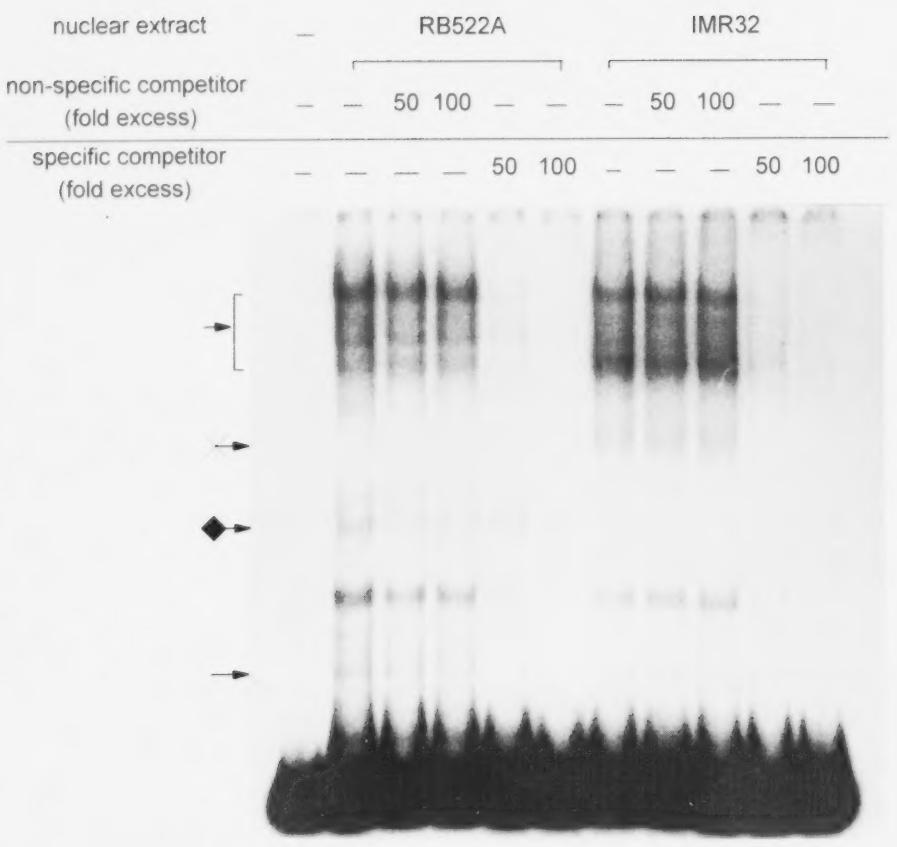

nuclear extract
RB522A
IMR32
non-specific competitor
(fold excess)
50
100
50
100
specific competitor
(fold excess)
50
100
50
100


2
-202 bp
-36 bp
-600 bp
-400 bp
-200 bp
+1
+200 bp
+400 bp

was also detected in the negative control lane. Interestingly, only two of four specific complexes were present in both RB522A and IMR32 nuclear extracts. One complex was specific to the RB522A nuclear extract, while another was specific to the IMR32 nuclear extract (Figure 16). Finally, EMSA using DNA fragment 4 (+163/+431) as the probe detected four DNA-protein complexes. Competition reactions indicated that all of them resulted from specific DNA-protein interaction (Figure 17).

DNase I footprinting assays were used to map binding sites for transcription factors involved in *DDX1* expression. DNA fragments 1-3 (Figure 13), which cover the region from -577 to +164 of *DDX1*, were used as probes in DNase I footprinting experiments. Each fragment was end-labelled on the coding strand and/or noncoding strand, incubated with nuclear extracts from RB522A or IMR32 cells and digested with DNase I. In negative control reactions, no nuclear extract was added. The digestion products were run on a sequencing gel. The sequence of the DNase I protected regions was determined by comparison with the sequence of the negative control sample and a G+A sequence ladder. DNA fragment 1 (-577/-203) was end-labelled at the coding and noncoding strands, respectively. DNase I footprinting of both coding and noncoding strands revealed six DNase I protected regions (Figure 18). Region III (-334 to -307 bp) was detected on both coding and noncoding strands. Regions I (-486 to -452 bp) and II (-428 to -392 bp) were on the noncoding strand. Regions IV (-296 to -274 bp), V (-267 to -239 bp) and VI (-220 to -203 bp) were on the coding strand. All these DNase I protected regions were present using both RB522A and IMR32 nuclear extracts. DNase I footprinting experiment with end-labelled noncoding strand of DNA fragment 2 (-202/-36) detected one DNase I protected region (VII) that was from -164 to -130 bp and was present using RB522A nuclear extracts but not IMR32 nuclear extracts (Figure 19). In addition, DNase I footprinting analysis of

Figure 16. Analysis of the interaction between nuclear proteins and DNA fragment 3 (-35/+163) of *DDX1* by EMSA. 2.5 μg of nuclear extract from either RB522A or IMR32 cells were incubated with 4 ng of labelled DNA fragment 3. Unlabelled DNA fragment 3 was used as specific competitor and the same DNA fragment as described in the legend to Figure 14 was used as non-specific competitor. Specific competitor and non-specific competitors were added in 50 and 100 fold excess compared to the labelled DNA fragment 3. Four specific retarded bands, representing nuclear proteins binding to sequence specific DNA, are indicated by arrows. The band that only appeared in IMR32 lanes is indicated by an arrow with "+". The band that only showed in RB522A lanes is indicated by an arrow with "◆". The diagram shows the location of DNA fragment 3 (-35/+163) in *DDX1*. The black box represents the first exon of *DDX1*. The transcription initiation site is at "+1".


nuclear extract
—
RB522A
IMR32
non-specific competitor
(fold excess)
— — 50 100 — — — 50 100 — —
specific competitor
(fold excess)
— — — — 50 100 — — — 50 100

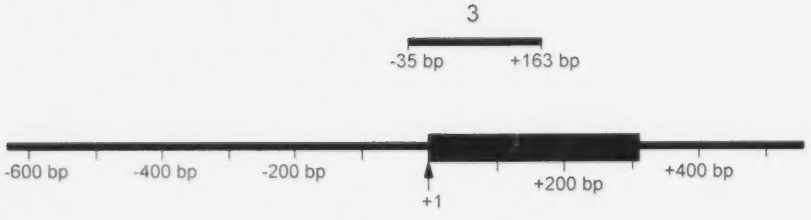

3
-35 bp
+163 bp
-600 bp
-400 bp
-200 bp
+1
+200 bp
+400 bp

Figure 17. Analysis of the interaction between nuclear proteins and DNA fragment 4 (+164/+431) of *DDX1* by EMSA. 1μg of nuclear extract from either RB522A or IMR32 cells was incubated with 4 ng of labelled DNA fragment 4. Unlabelled DNA fragment 4 was used as specific competitor and non-specific competitor used is as described in the legend to Figure 14. Specific and non-specific competitors were added in 50 and 100 fold excess compared to the labelled DNA fragment 4. Four specific retarded bands resulting from nuclear proteins binding to sequence specific DNA are indicated by arrows. The location of DNA fragment 4 (+164/+431) in *DDX1* is shown in the diagram. The black box represents the first exon of *DDX1* and the transcription initiation site is at "+1".

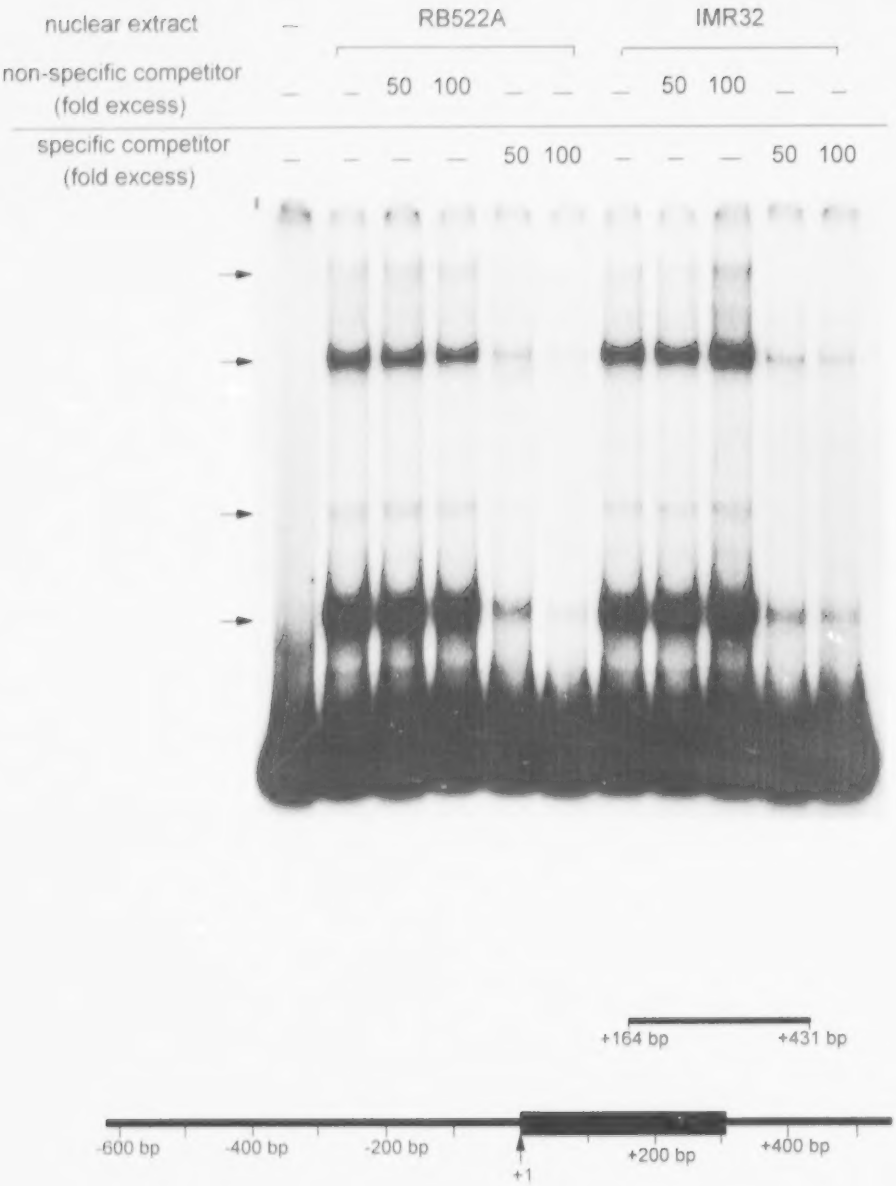

nuclear extract
–
RB522A
IMR32
non-specific competitor
(fold excess)
– – 50 100 – – – 50 100 – –
specific competitor
(fold excess)
– – – – 50 100 – – – 50 100
+164 bp
+431 bp
-600 bp
-400 bp
-200 bp
+1
+200 bp
+400 bp

Figure 18. DNase I footprint analysis of DNA fragment 1 (-577/-203) of *DDX1*. The fragment was end-labelled on the coding strand and noncoding strand respectively, incubated with either no nuclear extract (control lane) or 30 μg of nuclear extract from RB522A or IMR32 cells, and digested with DNase I. The numbers at the top of autoradiogram are the amounts of DNase I (μg/ml) used in each reaction mixture. Areas protected from DNase I digestion are bracketed with the corresponding nucleotide positions determined by a G+A sequencing ladder. The DNase I protected regions are marked by roman numerals. The diagram shows the location of the DNA fragment 1 in *DDX1*. The black box represents the first exon of *DDX1* and the transcription initiation site is at "+1".

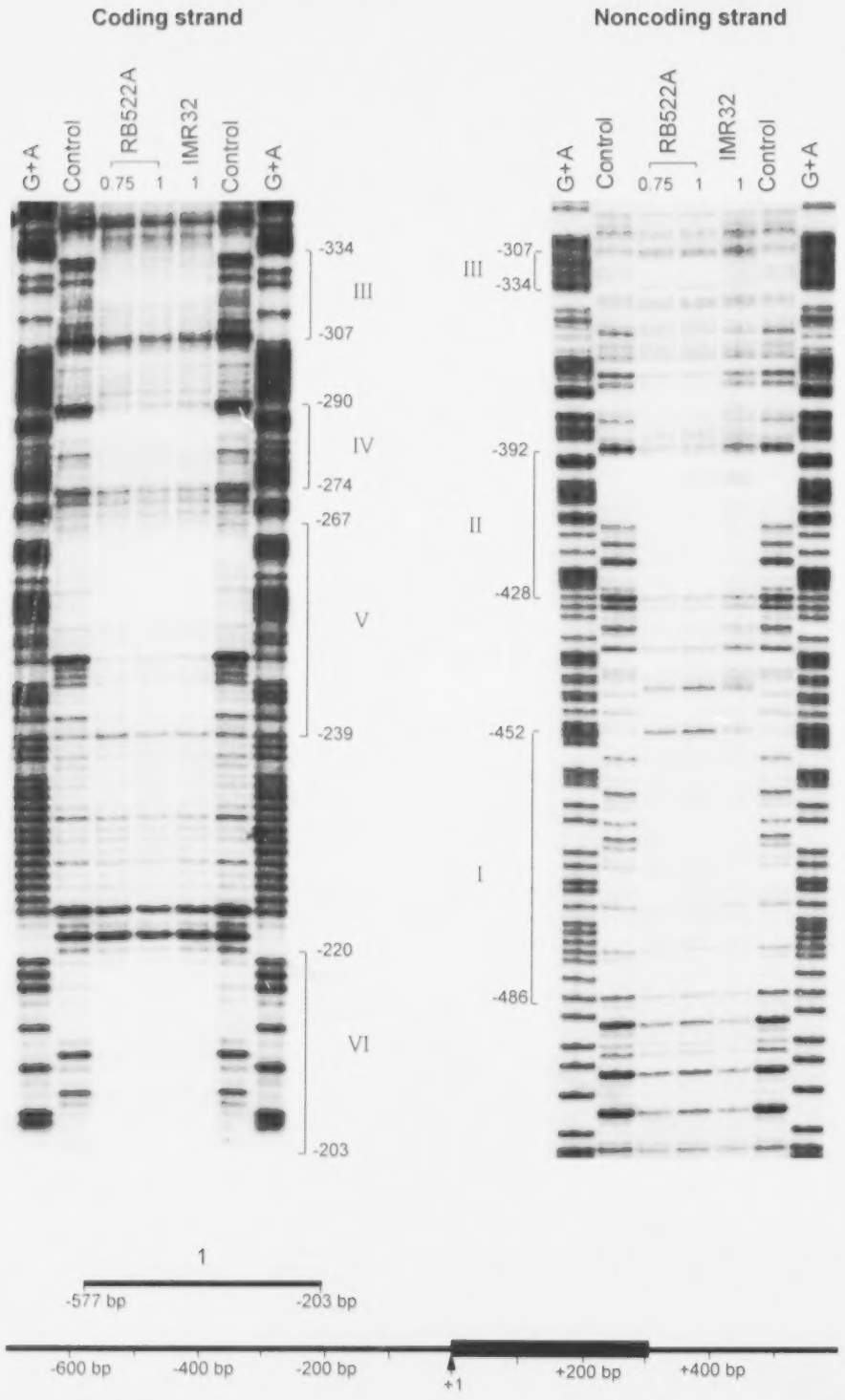
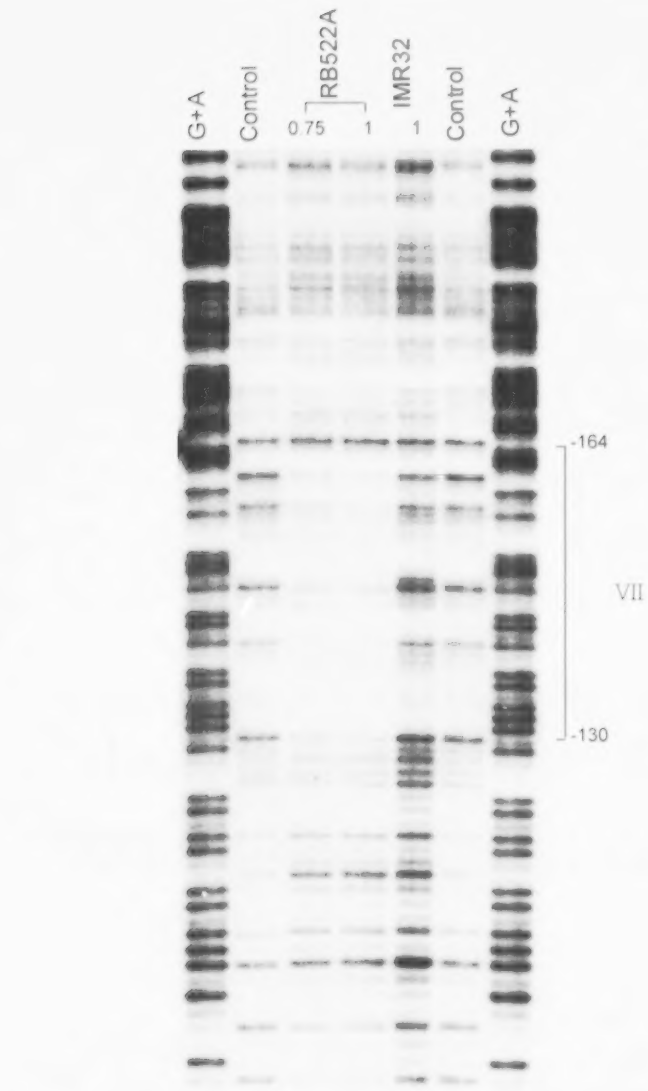
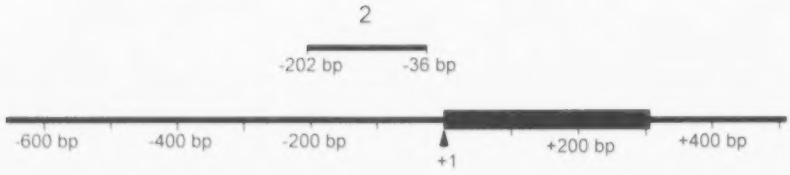

Coding strand
Noncoding strand
G+A
Control
RB522A
0.75
1
IMR32
1
Control
G+A
-334
-307
III
-290
-274
IV
-267
-239
V
-220
-203
VI
-307
-334
III
-392
-428
II
-452
-486
I
1
-577 bp
-203 bp
-600 bp
-400 bp
-200 bp
+1
+200 bp
+400 bp

Figure 19. DNase I footprint analysis of DNA fragment 2 (-202/-36) of *DDX1*. The fragment was end-labelled on the non-coding strand, incubated with either no nuclear extract (control lane) or 30 μg of nuclear extract from RB522A or IMR32 cells, and digested with DNase I. The numbers at the top of autoradiogram are the amounts of DNase I (μg/ml) used in each reaction mixture. The area protected from DNase I digestion is bracketed with corresponding nucleotide positions determined by a G+A sequencing ladder. The DNase I protected region is marked by roman numeral VII. The diagram shows the location of the DNA fragment 2 in *DDX1*. The black box represents the first exon of *DDX1* and the transcription initiation site is at "+1".

Noncoding strand



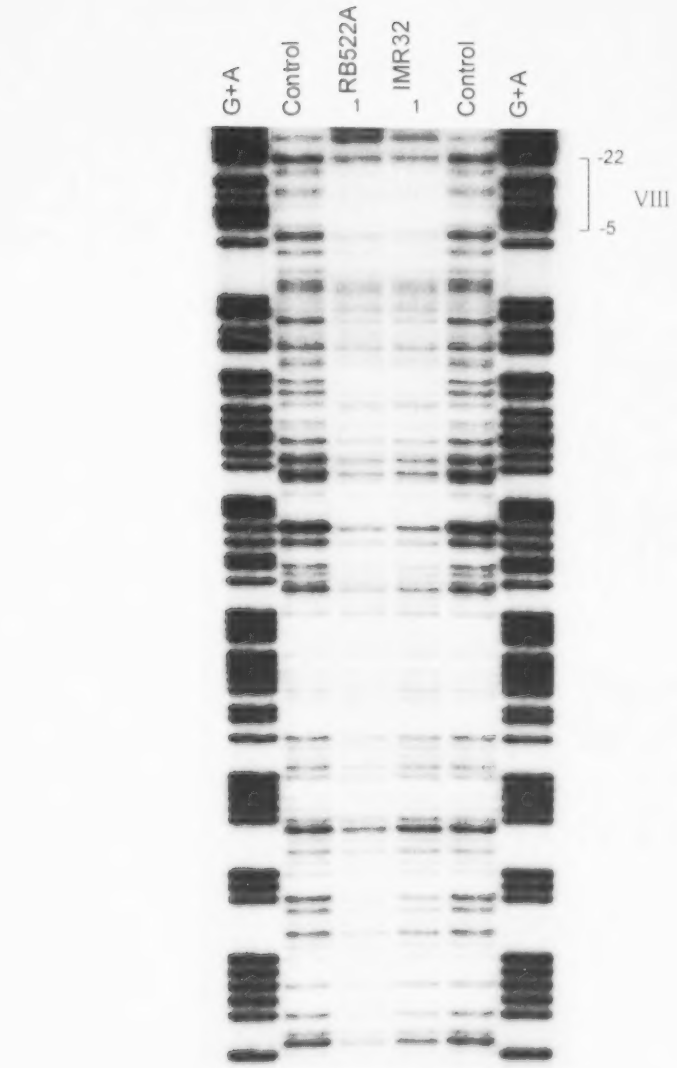


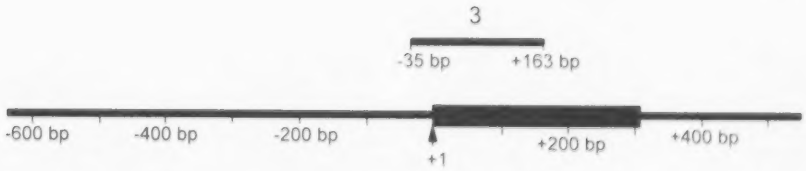
the coding strand of DNA fragment 3 (-35/+164) also revealed one DNase I protected region (VIII) from -32 to -5. This region was present using both RB522A and IMR32 nuclear extracts (Figure 20).

All the DNase I protected regions (I to VIII) obtained from DNase I footprinting experiments are shown in Figure 21. Dr. Rufus Day (Cross Cancer Institute) carried out a search of transcription factor binding sites for each of DNase I protected regions using "TransFactor 6", a program that locates the positions of consensus sites in a DNA sequence. The search revealed recognition sites for the following transcription factors: C/EBP, CF1, E2A, TBP, NF-IL6 and GATA-1,2,3. Binding sites for C/EBP (GCAAT) were found in DNase I protected regions I and IV. Binding sites for CF1 (AGATGG) and E2A (AGAGATG) were found in DNase I protected region III. In addition, binding sites for GATA-1,2,3 (AGATAA), TBP (TATAAA) and NF-IL6 (TAATGATTT) were found in DNase I protected regions II, V and VI, respectively. All these transcription factors have been implicated in cell growth and differentiation. These transcription factors and their consensus binding sites are summarized in Table 1. No binding sites for transcription factors identified so far were found in protected regions VII and VIII.

Figure 20. DNase I footprint analysis of DNA fragment 3 (-35/+164) of *DDX1*. The fragment was end-labelled on the coding strand, incubated with either no nuclear extract (control lane) or 30 μg of nuclear extract from RB522A or IMR32 cells, and digested with DNase I. The numbers at the top of autoradiogram are the amounts of DNase I (μg/ml) used in each reaction mixture. The area protected from DNase I digestion is bracketed with the corresponding nucleotide positions determined by a G+A sequencing ladder. The DNase I protected region is marked by roman numeral VIII. The diagram shows the location of the DNA fragment 3 in *DDX1*. The black box represents the first exon of *DDX1* and the transcription initiation site is at "+1".

Coding strand





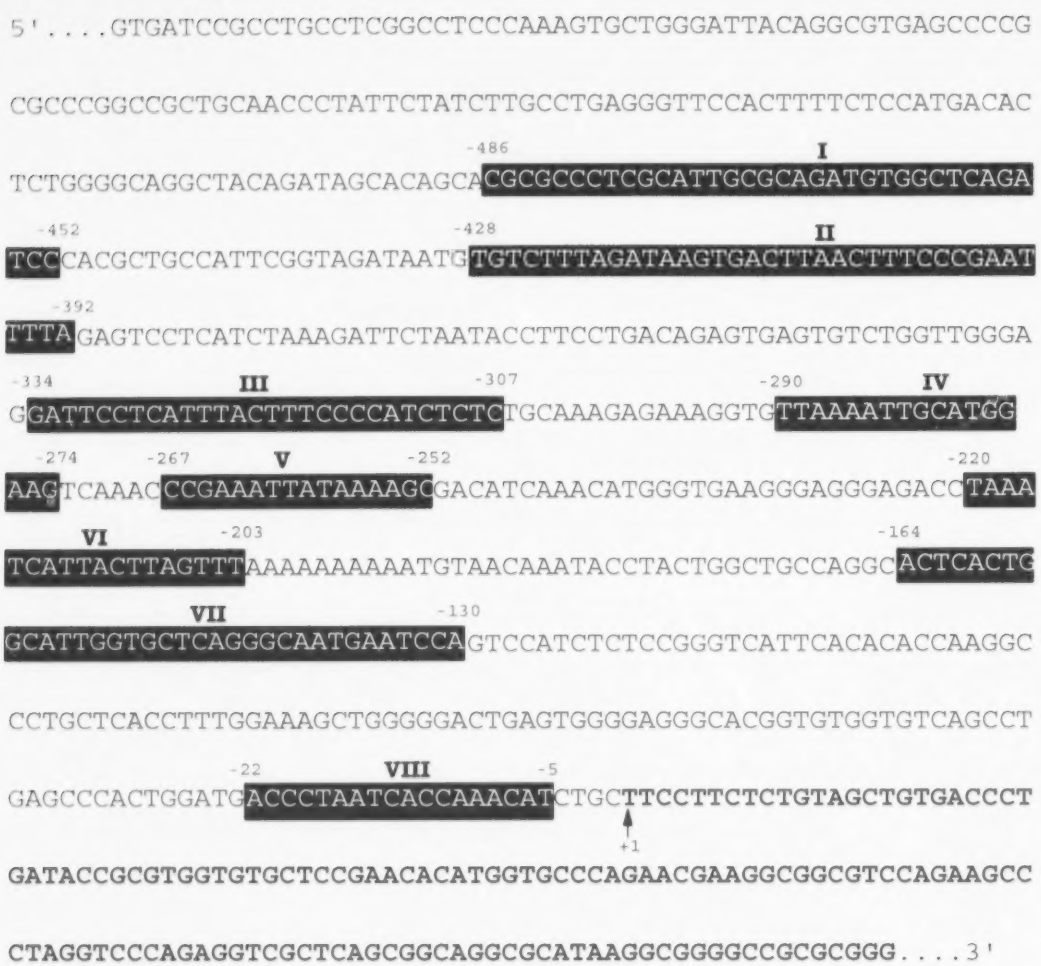


Figure 21. Sequence of DNase I protected regions I to VIII (shown in dark boxes). The sequence of the first exon of *DDX1* is shown in bold and the transcription initiation site is at "+1".

# of DNase I protected regions found in *DDX1*	I	II	III	IV	V	VI	VII	VIIII
Locations of the protected regions in the *DDX1 gene*	-486/-452	-428/-392	-334/-307	-290/-274	-267/-252	-220/-203	-164/-130	-22/-5
Transcription factors that may bind to those regions	C/EBP	GATA-1,2,3	CF1 E2A	C/EBP	TBP	NF-IL6	/	/
Consensus sites of transcription factors found in those regions	* GCAAT	AGATAA	*CF1: AGATGG *E2A:AGAGATG	* GCAAT	TATAAA	*TAATGATTT	/	/

Table 1. Results of searching transcription factor binding sites for DNase I protected regions in *DDX1* using "TransFactor 6" program. The direction of consensus site is from 5' to 3'.The consensus sites labelled by "*" are located on the noncoding strand.

CHAPTER V.

DISCUSSION AND FUTURE WORK

V. 1. Discussion

Loss of both copies of the *RB* gene is essential in tumorigenesis of retinoblastoma (RB). The product of the *RB* gene, pRB, plays a key role in regulation of cell growth and differentiation by modulating the cell cycle. On the other hand, pRB has also been shown to directly or indirectly affect expression of other genes by functioning as a transcription factor, which suggests that in RB, as the result of pRB inactivation, expression of some genes may be altered and these alterations in turn may have an implication in progression of RB. In an attempt to identify genes whose expression may be altered in RB tumour cells, we used a +/- screening strategy to isolate cDNAs preferentially expressed in RB cells compared to fetal tissues and a tumour that expresses the RB gene, colon carcinoma.

A human DEAD box gene, *DDX1*, was identified using this strategy (63). DEAD box proteins are a family of putative ATP-dependent RNA helicases. The members of this family have been implicated in many cellular functions including translation initiation, ribosomal assembly, RNA splicing, spermatogenesis and embryogenesis (65). *DDX1* has been shown to be amplified and over-expressed in two RB and several neuroblastoma cell lines amplified for *MYCN*, as well as in some primary neuroblastoma tumours amplified for *MYCN*, which suggest co-amplification of *DDX1* and *MYCN* in these cell lines and tumours (63, 100-103). The regulation of *DDX1* expression and the functions of DDX1 protein are unknown. In order to gain insight into the regulation of *DDX1* expression, the studies presented in this thesis were designed to characterize *DDX1* in RB and neuroblastoma cell lines by identifying the transcription initiation site of *DDX1*

and characterizing the regulatory elements involved in the regulation of *DDX1* transcription.

Using primer extension, three major and one minor transcription initiation sites of *DDX1* were identified in two different cell lines: the RB tumour cell line, RB522A, and the neuroblastoma cell line, BE(2)-C. Surprisingly, these transcription initiation sites were not found in another neuroblastoma cell line amplified for *DDX1*, IMR32. There are two possibilities to explain the absence of these transcription initiation sites in IMR32. First, the poly $(A)^+$ RNA isolated from IMR32 may be degraded due to long term storage or contamination with RNase during primer extension, which would result in no primer extension products. Second, the transcription initiation site of *DDX1* in IMR32 may be altered and this could result from *DDX1* rearrangement during amplification. It is believed that gene amplification involves unscheduled DNA synthesis followed by recombination (139). During amplification, bidirectional replication at an origin of replication generates a bubble that can undergo further rounds of unscheduled DNA replication, resulting in partially replicated duplexes (139). Recombination among these duplexes could generate abnormal structural arrangement of a gene. Analysis of the amplified domain containing *MYCN* in neuroblastoma cell lines showed that there is a low degree of sequence homogeneity among amplicons from different cell lines and a significant number of rearrangements in the amplicons within a cell line (140), which suggest that there is a different amplification process for the same gene in different cell lines and the structure of an amplified gene may be different in different cell lines.

Multiple regulatory elements are involved in transcription regulation of a eukaryotic gene, including promoter, enhancer and other elements. Promoter elements are proximal to the transcription initiation site. They have a role in determining the transcription initiation site and are involved in producing basal

levels of transcription. Most eukaryotic genes contain a TATA box located 25 to 30 bp upstream of transcription initiation site in their promoters. But the promoters of some genes, such as housekeeping genes, do not contain the TATA box and are known as TATA-less promoters (141). Enhancer elements are usually distally located from the transcription initiation site, either upstream or downstream or even within a gene. They increase the level of gene expression (142). Other regulatory elements are involved in the expression of tissue-specific genes or transcription regulation in response to particular signals (143). In the study presented in chapter IV of this thesis, we attempt to identify regulatory elements involved in regulation of DDX1 transcription. A ~2.1 kb 5' flanking region of DDX1 was obtained by genomic screening. In the sequence immediately upstream of the transcription initiation sites, no typical TATA box was found at its characteristic position. It is therefore likely that the promoter of DDX1 belongs to TATA-less promoters, which may be the reason for the presence of three major transcription initiation sites since transcription of TATA-less genes is known to be often initiated at several sites (144). It has been well documented that the sequences known as the initiator elements, which encompass the transcription initiation site, are able to initiate transcription in the TATA-less promoters (141, 145).

To search the possible locations of putative regulatory elements in *DDX1*, we used DNase I hypersensitivity assay. DNase I HS sites are known to be associated with the regulatory elements of transcribed gene (128) and have been mapped to promoters and enhancers (131). In the RB522A cell line, seven DNase I HS sites (HS I to VII) have been identified in the 5' flanking region, the first exon and the first intron of *DDX1* (Figure 12). One of them (HS IV) was mapped immediately upstream of the transcription initiation site and probably represents the promoter region that is responsible for initiation and basal level of

DDX1 transcription. Of the other six DNase I HS sites, four were mapped downstream of the transcription initiation site (HS III was within the first exon; HS I, II and V were within the first intron) and two (HS VI and VII) were mapped far upstream of the transcription initiation sites, at about -3.4 kb and -16.5 kb. These six DNase I HS sites are potential candidates for additional regulatory elements required for *DDX1* expression, such as enhancers, tissue-specific elements or other elements involved in the regulation of *DDX1* transcription. In the studies of transcription regulation of some viral and cellular genes, several regulatory elements located downstream of the transcription initiation site have been shown to have a function in the regulation of gene transcription (146-149). For example, the major elements involved in the regulation of *MYCN* transcription have been located within 1 kb immediately upstream of the transcription initiation site and also in the first exon and/or first intron of the *MYCN* gene (150).

In order to gain some information regarding the promoter region of *DDX1*, the ~2.1 kb 5' flanking region was analyzed for homology to other known eukaryotic promoters. A region from -315 to -272 bp in *DDX1* was found to be homologous to the *MYCN* promoter region from -214 to -170 bp. Study of the regulatory elements involved in *MYCN* expression using DNA transfection analysis has shown that one region about 220 bp upstream of the transcription start sites is responsible for basal level of *MYCN* expression (150). At the present time, it is unclear whether the *MYCN* promoter homology region in *DDX1* is involved in the regulation of *DDX1* transcription, but DNase I protected regions have been found in this region, which suggest that some transcription factors may bind to a specific DNA sequence in this region. It is known that DDX1 mRNA is expressed at higher levels in tissues of neuro-ectodermal origin, including brain, retina and spinal cord, than in other tissues (63). MYCN mRNA is also highly expressed in the central nervous system during development (151,

152). Since both *DDX1* and *MYCN* are expressed at higher levels in normal tissues of neuro-ectodermal origin compared to other tissues, the finding of *MYCN* promoter homology region in *DDX1* raises the possibility that the same transcription factor may be involved in the transcription regulation of both the *DDX1* and *MYCN* genes.

Based on the data obtained from DNase I hypersensitivity assay and eukaryotic promoter homology search of the ~2.1 kb 5' flanking region, we used EMSA and DNase I footprinting to further analyze the region from -577 to +431 bp of *DDX1* which includes the area with homology to the *MYCN* promoter. Specific DNA-protein interactions were observed in this region as determined by EMSA. Interestingly, four specific DNA-protein interactions were detected in the region from -35 to +163bp. Two of them were observed using both RB522A and IMR32 nuclear extracts, one was RB522A specific and the other was IMR32 specific. This result suggests that in RB522A and IMR32 nuclear extracts, there are different DNA binding proteins that recognize and bind to different sequences in this region. To further identify the areas of specific DNA-protein interactions, the region from -577 to +163 bp in *DDX1* was analyzed using DNase I footprinting. Eight DNase I protected regions (I to VIII, Figure 21) were identified and the protected region VII was RB522A specific, which again indicates that there are different binding proteins in RB522A and IMR32 nuclear extracts. All the DNase I protected regions identified to date are located upstream of the transcription initiation sites. These results suggest the presence of multiple regulatory elements located upstream of the *DDX1* gene.

Computer analysis of DNA-protein interaction sites identified by DNase I footprinting has revealed the presence of binding sites for several transcription factors: C/EBP, GATA-1,2,3, CF1, E2A, TBP, and NF-IL6 (Table 1). These transcription factors have been implicated in the regulation of gene transcription,

cell growth and differentiation. C/EBP (CCAAT/enhancer-binding protein) represents a family of transcription factors (153). The proteins of this family contain three structural components that include a C-terminal leucine-zipper, a basic DNA-binding region and a N-terminal transactivating region. Expression of C/EBP is elevated in adipocytes, hepatocytes, myeloid and B cells (154), which suggests that these proteins are involved in tissue-specific expression. Differentiation studies of cultured adipocytes (3T3-L1) showed that C/EBPα is required for preadipocyte differentiation (155). GATA-1,2,3 are erythrocyte-specific transcription factors (156, 157). They bind to the same consensus sequence (AGATAA) that is found in the regulatory regions of erythroid-specific genes (158, 159). Study of erythroid differentiation in chimaeric mice demonstrated that GATA-1 is required for the normal differentiation of erythroid cells (160). Transfection analysis showed that GATA-3 can activate transcription of a reporter gene through a GATA binding site identified within the human T-cell receptor δ gene, suggesting that GATA-3 plays a role in T-cell specific transcription regulation (161). E2A transcription factor is known to bind to the E2 motif [(G/A)CAG(G/A or C/T)TG] that is present in the enhancers of several genes, including the immunoglobulin heavy- and light-chain gene, the insulin gene and muscle creatine kinase gene (162-164). The function of E2A is not very clear. Some studies suggest that it may be involved in muscle differentiation by activation of muscle creatine kinase gene through interaction with MyoD (a factor that plays a role in skeletal muscle specific gene regulation) and Id (a negative regulatory of helix-loop-helix DNA binding protein) (165-166). CF1 (common factor 1) has been shown to bind to the *MYCC* promoter, the skeletal α -actin promoter and the immunoglobulin heavy-chain enhancer (IgHE) (167). The function of CF1 site in *MYCC* is unknown. Transfection analysis show that the binding site of CF1 in the skeletal α-actin promoter is essential for promoter

function (168) and that CF1 site in the IgHE appears to function as a positive activator (169). TBP (TATA-binding protein) is one of the most highly conserved eukaryotic proteins, which plays a critical role in initiation of transcription by forming a complex with TAFs (TBP-associated factors) and direct binding to the TATA box (170). The TBP binding site found in *DDX1* is located at -260 to -255 bp. We believe that it does not represent a true TBP binding site because it is located too far upstream of the transcription initiation site. Transcription factor NF-IL6 belongs to the C/EBP family, which specifically binds to a responsive element in the IL-6 (interleukin-6) gene (171). In addition, it was also shown to bind to the transcription regulatory regions found in various acute-phase protein genes, such as TNF (172). Because IL-6 and acute-proteins play important protective roles in the host defense against tissue damage and infections, it was suggested that NF-IL6 may be involved in the regulation of acute-phase reaction and inflammation (171). Binding sites for all the transcription factors described above have been found in the DNase I protected regions identified in *DDX1*. It is not yet clear whether they do actually bind to these regions *in vivo*, but it will be interesting to study their possible effects on *DDX1* expression.

V. 2. Future work

The studies of human DEAD box gene, *DDX1*, in RB and neuroblastoma cell lines are presented in this thesis.

We have identified three major transcription initiation sites in *DDX1*. In addition, the regulatory elements involved in the regulation of *DDX1* transcription were studied using DNase I hypersensitivity assay, EMSA and DNase I footprinting. DNase I HS sites and specific DNA-protein interactions were detected in the 5' flanking region, as well as within the first exon and the first intron of *DDX1*, which suggests that these regions contain regulatory elements.

Eight DNase I protected regions representing the sites of DNA-protein interactions have been identified in the region from -577 to -5 bp of *DDX1* and a search for transcription factor binding sites revealed the presence of binding sites for several known transcription factors, suggesting that these transcription factors may be involved in the regulation of *DDX1* transcription. Furthermore, one region from -315 to -272 bp in *DDX1* was identified which has homology to the *MYCN* promoter region from -214 to -170 bp. The latter region has been shown to result in a basal level of *MYCN* transcription. This suggests the possibility that the same transcription factor may be involved in transcription regulation of both *DDX1* and *MYCN*. Finally, the results from EMSA and DNase I footprinting indicate that there are different binding proteins in RB522A and IMR32, which recognize different DNA sequences in *DDX1*.

Future studies will be carried out to further identify the regulatory elements involved in the regulation of *DDX1* transcription. To eliminate the possibility that degradation of poly $(A)^+$ RNA isolated from IMR32 is the reason why no transcription initiation sites were detected in this cell line, the primer extension will be repeated using freshly prepared poly $(A)^+$ RNA and adding RNase inhibitor to the reaction to prevent degradation of the poly $(A)^+$ RNA. DNA transfection assay will be used to analyze the regions containing putative regulatory elements to determine their biological function in the regulation of *DDX1* transcription. Our strategy will be to first introduce a 2.1 kb of 5' flanking DNA into pCAT vector and to progressively study deletions of this fragment to determine when a decrease or increase in CAT activity is obtained. The same strategy will also be used to investigate other regions that appear to contain regulatory elements. Since the level of CAT activity reflects the activity of the regulatory elements located in the DNA fragment inserted, this strategy will identify the DNA elements involved in the regulation of *DDX1* transcription. In

addition, to further analyze the binding sites of DNA-protein interactions, oligonucleotides will be prepared based on the sequences of binding sites and used as probes in EMSA and DNase I footprinting to determine their potential functions as regulatory elements.

All these studies will contribute to a better understanding of how the expression of *DDX1* is regulated.

REFERENCE

1. Knudson, A. G. (1971). Mutation and cancer: statistical study of retinoblastoma. Proc. Natl. Acad. Sci. USA. 68, 820-823

2. Francke, U. and Kung, F. (1976). Sporadic bilateral retinoblastoma and 13q- chromosomal deletion. Med. Pediatr. Oncol. 2, 379-385

3. Knudson, A. G., Meadows, A. T., Nichols, W. W. and Hill, R. (1976). Chromosomal deletion and retinoblastoma. N. Engl. J. Med. 295, 1120-1123

4. Benedict, W. F., Murphree, A. L., Banerjee, A., Spina, C. A., Sparkes, M. C. and Sparkes, R. S. (1983). Patients with 13 chromosome deletion: evidence that the retinoblastoma gene is recessive cancer gene. Science 219, 973-975

5. Godbout, R., Dryja, T. P., Squire, J., Gallie, B. L. and Phillips, R. A. (1983). Somatic inactivation of genes on chromosome 13 is a common event in retinoblastoma. Nature 304, 451-453

6. Cavenee, W. K., Dryja, T. P., Phillips, R. A., Benedict, W. F., Godbout, R., Gallie, B. L., Murphree, A. L., Strong, L. C. and White, R. L. (1983). Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. Nature 305, 779-784

7. Dryja, T. P., Cavenee, W., White, R., Rapaport, J. M., Petersen, R., Albert, D. M. and Bruns, G. A. (1984). Homozygosity of chromosome 13 in retinoblastoma. N. Engl. J. Med. 310, 550-553

8. Fung, T. Y-K. T., Murphree, A. L., T'ang, A., Qian, J., Hinrichs, S. H. and Benedict, W. F. (1987). Structural evidence for the authenticity of the human retinoblastoma gene. Science 236, 1657-1661

9. Dunn, J. M., Phillips, R. A., Becker, A. J. and Gallie, B. L. (1988). Identification of germline and somatic mutations affecting the retinoblastoma gene. Science 241,1797-1800

10. Friend, S. H., Benards, R., Rogelj, S., Weinberg, R. A., Rapaport, J. M., Albert, D. M. and Dryja, T. P. (1986). A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. Nature 323, 643-646

11. Lee, W.H., Brookstein, R., Hong, F., Young, L. J., Shew, J. Y. and Lee, E. Y-H. P. (1987). Human retinoblastoma susceptibility gene: cloning, identification, and sequence. Science 235, 1394-1399

12. Lee, W. H., Shew, J. Y., Hong, F. D., Sery, T. W., Donoso, L. A., Young, L. J., Bookstein, R. and Lee, E. Y-H. P. (1987). The retinoblastoma susceptibility gene encodes a nuclear phosphoprotein associated with DNA binding activity. Nature 329, 642-645

13. Buchkovich, K., Duffy, L. A. and Harlow, F. (1989). The retinoblastoma protein is phosphorylated during specific phases of the cell cycle. Cell 58, 1097-1105

14. Chen, P. L., Scully, P., Shew, J. Y., Wang, J. Y. J. and Lee, W. H. (1989). Phosphorylation of the retinoblastoma gene product is modulated during the cell cycle and cellular differentiation. Cell 58, 1193-1198

15. Hansen, M. F., Koufos, A., Gallie, B. L., Phillips, R. A., Fodstad, O., Brogger, A., Gedde-Dahl, T. and Cavenee, W. K. (1985). Osteosarcoma and retinoblastoma: a shared chromosomal mechanism revealing recessive predisposition. Proc. Natl. Acad. Sci. USA. 82, 6216-6220

16. Harbour, J. W., Lai, S. L., Whang-Peng, J., Gazdar, A. F., Minna, J. D. and Kaye, F. J. (1988). Abnormalities in structure and expression of the human retinoblastoma gene in SCLC. Science 241, 353-357

17. Lee, E. Y-H. P., To, H., Shew, J. Y., Bookstein, R., Scully, P. and Lee, W. H. (1988). Inactivation of the retinoblastoma susceptibility gene in human breast cancers. Science 241, 218-221

18. Varley, J. M., Armour, J., Swallow, J. E., Jeffreys, A. J., Ponder, B. A. J., T'Ang, A., Fung, Y. K., Brammar, W. J. and Walker, R. A. (1989). The retinoblastoma gene is frequently altered leading to loss of expression in primary breast tumours. Oncogene 4, 725-729

19. Takahashi, R., Hashimoto, T., Xu, H. J., Hu, S. X., Matsui, T., Miki, T., Bigo-Marshall, H., Aaronson, S. A. and Benedict, W. F. (1991). The retinoblastoma gene functions as a growth and tumour suppressor in human bladder carcinoma cells. Proc. Natl. Acad. Sci. USA. 88, 5257-5261

20. Sarkar, F. H., Sakr, W., Li, Y. W., Macoska, J., Ball, D. E. and Crissman, J. D. (1992). Analysis of retinoblastoma (RB) gene deletion in human prostatic carcinoma. Prostate 21, 145-152

21. Chen, Y. C., Chen, P. J., Yeh, S. H., Tien, H. F., Wang, C. H., Tang, J. L. and Hong, R. L. (1990). Deletion of the human retinoblastoma gene in primary leukemias. Blood 76, 2060-2064

22. DeCaprio, J. A., Furukawa, Y., Ajchenbaum, F., Griffin, J. D. and Livingston, D. M. (1992). The retinoblastoma-susceptibility gene product becomes phosphorylated in multiple stages during cell cycle entry and progression. Proc. Natl. Acad. USA. 89, 1795-1798

23. Mittnacht, S., Hinds, P.W., Dowdy, S. F. and Weinberg, R. A. (1991). Modulation of retinoblastoma protein activity during the cell cycle. Cold Spring Harb. Symp. Quant. Biol. 56, 197-209

24. Goodrich, D. W., Wang, N. P., Qian, Y. W., Lee, E. Y-H. P. and Lee, W. H. (1991). The retinoblastoma product regulates progression through the G1 phase of the cell cycle. Cell 67, 293-302

25. DeCaprio, J. A., Ludlow, J. W., Figge, J., Shew, J. Y., Huang, C. M., Lee, W. H., Marsilio, E., Paucha, E. and livingston, D. M. (1988). SV40 large tumour antigen forms a specific complex with the product of the retinoblastoma susceptibility gene. Cell 54, 275-283

26. Whyte, P., Buchkovich, K. J., Horowitz, J. M., Friend, S. H., Raybuck, M., Weinberg, R. A. and Harlow, E. (1988). Association between an oncogene and an anti-oncogene: the adenovirus E1A proteins bind to the retinoblastoma gene product. Nature 334, 124-129

27. Dyson, N., Howley, P. M., Münger, K. and Harlow, E. (1989). The human papilloma virus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product. Science 243, 934-937

28. Ludlow, J. W., Shen, J., Pipas, J. M., Livingston, D. M. and DeCaprio, J. A. (1990). The retinoblastoma susceptibility gene product undergoes cell cycle-dependent dephosphorylation and binding to and release from SV40 large T. Cell 60, 387-396

29. Ludlow, J. W., DeCaprio, J. A., Huang, C. M., Lee, W. H., Paucha, E. and Livingston, D. M. (1989). SV40 large T antigen binds preferentially to an underphosphorylated member of the retinoblastoma susceptibility gene product family. Cell 56, 57-65

30. Ewen, M,. E., Ludlow, J. W., Marsilio, E., DaCaprio, J. A., Millikan, R. C., Cheng, S. H., Paucha, E. and Livingston, D. M. (1989). An N-terminal transformation-governing sequence of SV40 large T antigen contributes to the binding of both $p110^{Rb}$ and a second cellular protein, p120. Cell 58, 257-267

31. DeCaprio, J. A., Ludlow, J. W., Lynch, D., Furukawa, Y., Griffin, J., Piwnica-Worms, H., Huang, C. M. and Livingston, D. M. (1989). The product of the retinoblastoma susceptibility gene has properties of a cell cycle regulatory element. Cell 58, 1085-1095

32. Chellappan, S. P., Hiebert, S., Mudryj, M., Horowitz, J. M. and Nevins, J. R. (1991). The E2F transcription factor is cellular target for the RB protein. Cell 65, 1053-1061

33. Nevins J. R. (1992). E2F: a link between the RB tumour suppressor protein and viral oncoproteins. Science 258, 424-429

34. Fattaey, A. R., Helin, K., Dembski, M. S., Dyson, N., Harlow, E., Vuocolo, G. A., Hanobik, M. G., Haskell, K. M., Oliff, A. and Defeo-Jones, D. (1993). Characterization of the retinoblastoma binding proteins RBP1 and RBP2. Oncogene 8, 3149-3156

35. Durfee, T., Becherer, K., Chen, P. L., Yeh, S. H., Yang, Y., Kilburn, A. E., Lee, W. H. and Elledge, S. J. (1993). The retinoblastoma protein associates with the protein phosphotase type 1 catalytic subunit. Genes Dev. 7, 555-569

36. Rustgi, A. K., Dyson, N. and Bernards, R. (1991). Amino-terminal domains of *c-myc* and *N-myc* proteins mediate binding to the retinoblastoma gene product. Nature 352, 541-544

37. Blake, M. C. and Azizkham, J. C. (1989). Transcription factor E2F is required for efficient expression of the hamster dihydrofolate reductase gene *in vitro* and *in vivo*. Mol. Cell. Biol. 9, 4994-5002

38. Dou, Q. P., Markell, P. J. and Pardee, A. B. (1992). Thymidine kinase transcription is regulated at G1/S phase by a complex that contains retinoblastoma-like protein and a cdc2 kinase. Proc. Natl. Acad. USA. 89, 3256-3260

39. Pearson, B. E., Nasheuer, H. P. and Wang, T. S. F. (1991). Human DNA polymerase α gene: sequence controlling expression in cycling and serum stimulated cells. Mol. Cell. Biol. 11, 2081-2095

40. Johnson, D. G., Schwarz, J. K., Cress, W. D. and Nevins, J. R. (1993). Expression of transcription factor E2F-1 induces quiescent cells to enter S phase. Nature 365, 349-352

41. Bandara, L. R. and La Thangue, N. B. (1991). Adenovirus E1a prevents the retinoblastoma gene product from complexing with a cellular transcription factor. Nature 351, 494-497

42. Hu, Q., Dyson, N. and Harlow, E. (1990). The regions of the retinoblastoma protein needed for binding to adenovirus E1A or SV40 large T antigen are common sites for mutations. EMBO. J. 9, 1147-1155

43. Kaelin, W. G. Jr., Ewen, M. E. and Livingston, D. M. (1990). Definition of the minimal simian virus 40 large T antigen- and adenovirus E1A-binding domain in the retinoblastoma gene product. Mol. Cell. Biol. 10, 3761-3769

44. Hiebert, S. W. (1993). Regions of the retinoblastoma gene product required for its interaction with the E2F transcription factor are necessary for E2 promoter repression and pRB-mediated growth suppression. Mol. Cell. Biol. 13, 3384-3391

45. Defeo-Jones, D., Huang, P.S., Jones, R. E., Haskell, K. M., Vuocolo, G. A., Hanobik, M. G., Huber, H. E. and Oliff, A. (1991). Cloning of cDNAs for cellular proteins that bind to the retinoblastoma gene product. Nature 352, 251-254

46. Helin, K., Lees, J. A., Vidal, M., Dyson, N., Harlow, E. and Fattaey, A. (1992). A cDNA encoding a pRB-binding protein with properties of the transcription factor E2F. Cell 70, 337-350

47. Kaelin, W. G. Jr., Krek, W., Sellers, W. R., DeCaporio, J. A., Ajchenbaum, F., Fuchs, C. S., Chittenden, T., Li, Y., Farnham, P. J., Blanar, M. A., Livingston, D. M. and Flemington, E. K. (1992). Expression cloning of a cDNA encoding a retinoblastoma-binding protein with E2F-like properties. Cell 70, 351-364

48. Shan, B., Zhu, X., Chen, P. L., Durfee, T., Yang, Y., Sharp, D. and Lee, W. H. (1992). Molecular cloning of cellular genes encoding retinoblastoma-associated proteins: identification of a gene with properties of the transcription facter E2F. Mol. Cell. Biol. 12, 5620-5631

49. Cobrinik, D., Dowdy, S. F., Hinds, P. W., Mittnacht, S. and Weinberg, R. A. (1992). The retinoblastoma protein and the regulation of cell cycling. Trends Biochem. Sci. 17, 312-315

50. Hamel, P. A., Phillips, R. A., Muncaster, M. and Gallie, B. L. (1993). Speculations on the roles of RB1 in tissues-specific differentiation, tumour initiation, and tumour progression. FASEB J. 7, 846-854

51. Slack, R. S., Hamel, P. A., Bladon, T. S., Gill R. M. and McBurney, M. W. (1993). Regulated expression of the retinoblastoma gene in differentiating embryonal carcinoma ceils. Oncogene 8, 1585-1591

52. Robbins, P. D., Horowitz, J. M. and Mulligan, R. C. (1990). Negative regulation of human *c-fos* expression by the retinoblastoma gene product. Nature 346, 668-671

53. Holt, J. T., Gopal, T. V., Moulton, A. D. and Nienhuis, A. W. (1986). Inducible production of *c-fos* antisense RNA inhibits 3T3 cell proliferation. Proc. Natl. Acad. USA. 83, 4794-4798

54. Riabowol, K. T., Vosatka, R. J., Ziff, E. B., Lamb, N. J. and Feramisco, J. R. (1988). Microinjection of *fos*-specific antibodies blocks DNA synthesis in fibroblast cells. Mol. Cell. Biol. 8, 1670-1676

55. Massagué, J., Cheifetz, S., Endo, T. and Bernardo, N. G. (1986). Type β transforming growth factor is an inhibitor of myogenic differentiation. Proc. Natl. Acad. Sci. USA. 83, 8206-8210

56. Masui, T., Wakefield, L. M., Lechner, J. F., laveck, M. A., Sporn, M. B. and Harris, C. C. (1986). Type β transforming growth factor is the primary differentiation inducing serum factor for normal human bronchial epithelial cells. Proc. Natl. Acad. Sci. USA. 83, 2438-2442

57. Roberts, A. B., Flanders, K. C., Heine, U. I., Jakowlew, S., Kondaiah, P., Kim, S. J. and Sporn, M. B. (1990). Transforming growth factor-beta: multifunctional regulator of differentiation and development. Philosophical Transactions of the Royal Society of London-Series B: Biological Sciences 327, 145-154

58. Kim, S. J., Lee, H. D., Robbins, P. D., Busam, K., Sporn, M. B. and Roberts A. B. (1991). Regulation of transforming growth factor β1 gene expression by the product of the retinoblastoma-susceptibility gene. Proc. Natl. Acad. Sci. USA. 88, 3052-3056

59. Kim, S. J., Wagner, S., Liu, F., O'Reilly, M. A., Robbins, P. D. and Green, M. R. (1992). Retinoblastoma gene product activates expression of the human *TGF-β2* gene through transcription factor ATF-2. Nature 358, 331-335

60. Marcu, K. B., Bossone, S. A. and Patel, A. J. (1992). Myc function and regulation. Annu. Rev. Biochem. 68, 809-860

61. Pietenpol, J. A., Stein, R. W., Moran, E., Yaciuk, P., Schlegel, R., Lyons, R. M., Pittelkow, M. R., Münger, K., Howley, P. M. and Moses, H. L. (1990). *TGF-β1* inhibition of *c-myc* transcription and growth in keratinocytes is abrogated by viral transforming proteins with pRB binding domains. Cell 61, 777-785

62. Pietenpol, J. A., Münger, K., Howley, P. M., Stein, R. W. and Moses, H. L. (1991). Factor-binding element in the human *c-myc* promoter involved in transcriptional regulation by transforming growth factor *β1* and by the retinoblastoma gene product. Proc. Natl. Acad. Sci. USA. 88, 10227-10231

63. Godbout, R. and Squire, J. (1993). Amplification of a DEAD box protein gene in retinoblastoma cell lines. Proc. Natl. Acad. Sci. USA. 90, 7578-7582

64. Linder, P., Lasko, P. F., Leroy, P., Nielsen, P. J. and Nishi, K. (1989). Birth of the D-E-A-D box . Nature 337, 121-122

65. Schmid, S. R. and Linder, P. (1992). D-E-A-D protein family of putative RNA helicase. Mol. Micro. 6, 283-292

66. Pause, A. and Sonenberg, N. (1992). Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A. EMBO J. 11, 2643-2654

67. Schmid, S. R. and Linder, P. (1991). Translation initiation factor 4A from *Saccharomyces cerevisiae*: analysis of residues conserved in the D-E-A-D family of RNA helicases. Mol. Cell. Biol. 11, 3463-3471

68. Walker, J. E., Saraste, M., Runswick, M. J. and Gay, N. J. (1982). Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. 1, 945-951

69. Rozen, F., Pelletier, J., Trachsel, H. and Sonenberg, N. (1989). A lysine substitution in the ATP-binding site of eukaryotic initiation factor 4A abrogates nucleotide-binding activity. Mol. Cell. Biol. 9, 4061-4063

70. Wassarman, D. A. and Steitz, J. A. (1991). Alive with DEAD proteins. Nature 349, 463-464

71. Hodgman, T. C. (1988). A new superfamily of replicative proteins. Nature 333, 22-23

72. Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P. and Blinov, V. M. (1989). Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. Nucleic Acids Res. 17, 4713-4730

73. Pause, A., Méthot, N. and Sonenberg, N. (1993). The HRIGRXXR region of the DEAD box RNA helicase in eukaryotic translation initiation factor 4A is required for RNA binding and ATP hydrolysis. Mol. Cell. Biol. 13, 6789-6793

74. Rozen, F., Edery, I., Meerovitch, K., Dever, T. E., Merrick, W. C. and Sonenberg, N. (1990). Bidirectional RNA helicase activity of eukaryotic translation initiation factor 4A and 4F. Mol. Cell. Biol. 10, 1134-1144

75. Blum, S., Schmid, S. R., Pause, A., Buser, P., Linder, P., Sonenberg, N. and Trachsel, H. (1992). ATP hydrolysis by initiation factor 4A is required for translation initiation in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA. 89, 7664-7668

76. Hirling, H., Scheffner, M., Restle, T. and Stahl H. (1989). RNA helicase activity associated with the human p68 protein. Nature 339, 562-564

77. Iggo, R. D. and Lane, D. P. (1989). Nuclear protein p68 is an RNA-dependent ATPase. EMBO J. 8, 1827-1831

78. Iggo, R. D., Jamieson, D. J., Macneill, S. A., Southgate, J., Mcpheat, J. and Lane, D. P. (1991). p68 RNA helicase: identification of a nucleolar form and cloning of related genes containing a conserved intron in yeasts. Mol. Cell. Biol. 11, 1326-1333

79. Edery, I., Pelletier, J. and Sonenberg, N. (1987). Role of eukaryotic RNA cap-binding protein in regulation of translation, pp. 335-366. In J. Ilan (ed), Translational regulation of gene expression. Plenum Publishing Crop., New York.

80. Rhoads, R. E. (1988). Cap-recognition and the entry of mRNA into the protein synthesis initiation cycle. Trends Biochem. Sci. 13, 52-56

81. Sonenberg, N. (1988). Cap binding proteins of eukaryotic messenger RNA: functions in initiation and control of translation. Prog. Nucleic Acids Res. Mol. Biol. 35, 174-207

82. Nielsen, P. J., McMaster, G. K. and Trachsel, H. (1985). Cloning of eukaryotic protein synthesis initiation factor genes: isolation and characterization of cDNA clones encoding factor eIF-4A. Nucleic Acids Res. 13, 6867-6880

83. Linder, P. and Slonimski, P. P. (1989). An essential yeast protein, encoded by duplicated genes *TIF1* and *TIF2* and homologous to the mammalian translation initiation factor eIF-4A, can suppress a mitochondrial missense mutation. Proc. Natl. Acad. Sci. USA. 86, 2286-2290

84. Dalbadie-McFarland, G. and Abelson, J. (1990). PRP5: a helicase-like protein required for mRNA splicing in yeast. Proc. Natl. Acad. Sci. USA. 87, 4236-4240

85. Strauss, E. J. and Guthrie, C. (1994). PRP28, a 'DEAD-box' protein, is required for the first step of mRNA splicing *in vitro*. Nucleic Acids Res. 22, 3187-3193

86. Jamieson, D. J., Rahe, B., Pringle, J. and Beggs, J. D. (1991). A suppressor of a yeast splicing mutation (*prp 8-1*) encodes a putaive ATP-dependent RNA helicase. Nature 349, 715-717

87. Nishi, K., Morel-Deville, F., Hershey, J. W. B., Leighton, T. and Schnier, J. (1988). An eIF-4A-like protein is a suppressor of an *Escherichia coli* mutant defective in 50S ribosomal subunit assembly. Nature 336, 496-498

88. Leroy, P., Alzari, P., Sassoon, D., Wolgemuth, D. and Fellous, M. (1989). The protein encoded by murine male germ cell-specific transcript is a putative ATP-dependent RNA helicase. Cell 57, 549-559

89. Lasko, P. F. and Ashburner, M. (1988). The product of the *Drosophila* gene *vasa* is very similar to eukaryotic initiation factor-4A. Nature 335, 611-617

90. Hay, B., Jan, L. Y. and Jan, Y. N. (1988). A protein component of *Drosophila* polar granules is encoded by *vasa* and has extensive sequence similarity to ATP-dependent helicases. Cell 55, 577-587

91. Liang, L., Diehl-Jones, W. and Lasko, P. (1994). Localization of vasa protein to the *Drosophila* pole plasm is independent of its RNA-binding and helicase activities. Development 120, 1201-1211

92. Valoir, T. D., Tucker, M. A., Belikoff, E. J., Camp, L. A., Bolduc, C. and Beckingham, K. (1991). A second maternally expressed *Drosophila* gene encodes a putative RNA helicase of the "DEAD box" family. Proc. Natl. Acad. Sci. USA. 88, 2113-2117

93. Lane, D. P. and Hoeffler, W. K. (1980). SV40 large T shares an antigenic determinant with a cellular protein of molecular weight 68,000. Nature 288, 167-170

94. Iost, I. and Dreyfus, M. (1994). mRNAs can be stabilized by DEAD-box proteins. Nature 372, 193-196

95. Buelt, M. K., Glidden, B. J. and Storm, D. R. (1994). Regulation of p68 helicase by calmodulin and protein kinase C. J. Biol. Chem. 269, 29367-29370

96. Alexander, K. A., Wakim, B. T., Doyle, G. S., Walsh, K. A. and Storm, D. R. (1988). Identification and characterization of the camoduline-binding domain of neuromodulin, a neurospecific camodulin-binding protein. J. Biol. Chem. 263, 7544-7549

97. Liu, Y. C. and Storm, D. R, (1990). Regulation of free calmoduline levels by neuromodulin: neuron growth and regeneration. Trends Pharmacol. Sci. 11, 107-111

98. Takagi, T. and Cox, J. A. (1990). Primary structue of the target of calcium vector protein of amphioxus. J. Biol. Chem. 265, 19721-19727

99. Espreafico, E. M., Cheney, R. E., Matteoli, M., Nascimento, A. A. C., DeCamilli, P. V., Larson, R. E. and Mooseker, M. S. (1992). Primary structure and cellular localization of chicken brain myosin-V (p190), an unconventional myosin with calmodulin light chains. J. Cell Biol. 119, 1541-1557

100. Squire, J. A., Thorner, P. S., Weitzman, S., Maggi, J. D., Dirks, P., Doyle, J., Hale, M. and Godbout, R. (1995). Co-amplification of *MYCN* and a DEAD box gene (*DDX1*) in primary neuroblastoma. Oncogene 10, 1417-1422

101. Manohar, C. F., Salwen, H. R., Brodeur, G. M. and Cohn, S. L. (1995). Co-amplification and concomitant high levels of expression of a DEAD box gene with *MYCN* in human neuroblastoma. Genes Chromosome Cancer 14, 196-203

102. Noguchi, T., Akiyama, K., Yokoyama, M., Kanda, N., Matsunaga, T. and Nishi, Y. (1996). Amplification of a DEAD box gene (*DDX1*) with the *MYCN*

gene in neuroblastoma as a result of cosegregation of sequences flanking the *MYCN* locus. Genes Chromosome Cancer 15, 129-133

103. George, R. E., Kenyon, R. M., McGuckin, A. G., Malcolm, A. J., Pearson, A. D. J. and Luneu, J. (1996). Investigation of co-amplification of the canadiate genes ornithine decarboxylase, ribonucleotide reductase, syndecan-1 and a DEAD box gene, *DDX1*, with *N-myc* in neurblastoma. Oncogene 12, 1583-1587

104. Kanda, N., Schreck, R., Alt, F., Bruns, G., Baltimore, D. and Latt, S. (1983). Isolation of amplified DNA sequences from IMR32 human neuroblastoma cells: Facilitation by fluorescence-activated flow sorting of metaphase chromosomes. Proc. Natl. Acad. Sci. USA. 80, 4069-4073

105. Schwab, M., Alitalo, K., Klempnauer, K. H., Varmus, H. E., Bishop, J. M., Gilbert, F., Brodeur, G., Goldstein, M. and Trent, J. (1983). Amplified DNA with limited homology to *myc* cellular oncogene is shared by human neuroblastoma cell lines and a neuroblastoma tumour. Nature 305, 245-248

106. Godbout, R., Hale, M. and Bisgrove, D. (1994). A human DEAD box protein with partial homology to heterogenous nuclear ribonucleoprotein U. Gene 138, 243-245

107. Dreyfuss, G., Swanson, M. S. and Pinol-Roma, S. (1988). Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. Trends Biochem. Sci. 13, 86-91

108. Kiledjian, M. and Dreyfuss, G. (1992). Priamry structure and binding activity of the hn RNA U protein: binding RNA through RGG box. EMBO J. 11, 2655-2664

109. Blackwood, E. M. and Eisenman, R. N. (1991). Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with *MYC*. Science 251, 1211-1217

110. Cole, M. D. (1991). Myc meets its Max. Cell 65, 715-716

111. Sambrook, J., Fritsch, E. F. and Maniatis. T. (1989). Molecular cloning -- a laboratory manual. Second edition. Cold Spring Harbor Laboratory Press. pp. 2.60-2.81

112. Davis, L. G., Dibner, M. D. and Battery, J. F. (1986). Basic methods in molecular biology. (Elsevier, New York).

113. Sambrook, J., Fritsch, E. F. and Maniatis, T (1989). Molecular cloning -- a laboratory manual. Second edition. Cold Spring Harbor Laboratory Press. pp. 10.6-10.10

114. Garabedian, M. J., LaBaer, J., Liu, W. H. and Thomas, J. R. Gene transcription: a practical approach. (1993). IRL Press. Oxford university Press. pp. 244-247

115. Garabedian, M. J., LaBaer, J., Liu, W. H. and Thomas, J. R. Gene transcription: a practical approach. (1993). IRL Press. Oxford university Press. pp. 262-264

116. Sanger, F., Nicklen, S. and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA. 74, 5463-5467

117. Mierendorf, R. C. and Pfeffer, D. (1987). Direct sequencing of denatured plasmid DNA. Methods Enzymol. 152, 556-562

118. Henikoff, S. (1987). Unidirectional digestion with exonuclease III in DNA sequence analysis. Methods Enzymol. 155, 156-165

119. Devereux, J., Haeberli, P. and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. Nucleic Acides Res. 12, 387-395

120. Dölz, R. (1994). GCG: The analysis of RNA secondary structure. Meth. in Mol. Biol. 24, 159-166

121. Zuker, M. (1994). Prediction of RNA secondary structure by energy minimization. Meth. in Mol. Biol. 25, 267-296

122. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). Molecular cloning -- a laboratory manual. Second edition. Cold Spring Harbor Laboratory Press. pp. 9.34-9.58

123. Godbout, R., Bisgrove, D. A., Honoré, L. H. and Day III, R. S. (1993). Amplification of the gene encoding the α-subunit of the mitochondrial ATP synthase complex in a human retinoblastoma cell line. Gene 123, 195-201

124. Godbout, R. and Tilghman, S. M. (1988). Configuration of α-fetoprotein regulatory domain during development. Genes Dev. 2, 949-956

125. Roy, R. J., Grosselin, P. and Guérin, S. L. (1991). A short protocol for micro-purification of nuclear proteins from whole animal tissue. BioTechniques 11, 770-777

126. Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72, 248-254

127. Garabedian, M. J., LaBaer, J., Liu, W. H. and Thomas, J. (1992). Analysis of protein-DNA interactions. Gene Transcription: a practical approach. IRL Press at Oxford University Press. pp.243-256

128. Eissenberg, J. C., Cartwright, I. L., Thomas, G. H. and Elgin, S. C. R. (1985). Selected topics in chromatin structure. Annu. Rev. Genet. 19, 485-536

129. Wu, C. (1980). The 5' ends of *Drosophila* heat shock genes in chromatin are hypersensitive to DNase I. Nature 286, 854-860

130. Wu, C. (1984). Two protein-binding sites in chromatin implicated in the activation of heat-shock genes. Nature 309, 229-234

131. Elgin, S. C. R. (1988). The formation and function of DNase I hypersensitive sites in the process of gene activation. J. Biol. Chem. 263, 19259-19263

132. Galas, D. J. and Schmitz, A. (1978). DNase I footprinting: a simple method for the detection of protein-DNA binding specificity. Nucleic Acids Res. 5, 3157-3170

133. Fried, M. and Crothers, D. M. (1981). Equilibria and kinetics of *lac* repressor-operator interactions by polyacrylamide gel electrophoresis. Nucleic Acids Res. 9, 6505-6525

134. Garner, M. M. and Revzin, A. (1981). A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Esherichia coli* lactose operon regulatory system. Nucleic Acids Res. 9, 3047-3060

135. Sigman, D. S. and Chen, C. H. (1990). Chemical nucleases: new reagents in molecular biology. Annu. Rev. Biochem. 59, 207-236

136. Docherty, K. and Clark, A. R. (1992). The characterization of transcriptional regulatory sequences. Gene transcription: a practical approach. IRL Press at Oxford University Press. pp. 87-101

137. Stanton, L. W., Schwab, M. and Bishop, J. M. (1986). Nucleotide sequence of the human *N-myc* gene. Proc. Natl. Acad. Sci. USA. 83, 1772-1776

138. Stanton, L. W. and Bishop, J. M. (1987). Alternative processing of RNA transcribed from *NMYC*. Mol. Cell. Biol. 7, 4266-4272

139. Stark, G. R. and Wahl, G. M. (1984). Gene amplification. Annu. Rev. Biochem. 53, 447-491

140. Akiyama, K. and Nishi, Y. (1990). Cloning and physical mapping of DNA sequences encompassing a region in *N-myc* amplicons of a human neurblastoma cell line. Nucleic Acids Res. 19, 6887-6894

141. Weis, L. and Reinberg, D. (1992). Transcription by RNA polymerase II: initiator-directed formation of transcription-competent complexes. FASEB J. 6, 3300-3309

142. Muller, M. M., Gerster, T. and Schaffner, W. (1988). Enhancer sequences and the regulation of gene transcription. Euro. J. Bioch. 176, 485-495

143. Maniatis, T., Goodbourn, S. and Fischer, J. A. (1987). Regulation of inducible and tissue-specific gene expression. Science 236, 1237-1245

144. Dynan, W. S. and Tjian, R. (1983). The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. Cell 35, 79-87

145. Smale, S. T. and Baltimore, D. (1989). The "initiator" as a transcription control element. Cell 57, 103-113

146. Kashanchi, F., Duvall, J. F., Lindholm, P. F., Radonovich, M. F. and Brady, J. N. (1993). Sequences downstream of the RNA initiation site regulate human T-cell lymphotropic virus type I basal gene expression. J. Virol. 67, 2894-2902

147. Mautner, J., Joos, S., Werner, T., Erick, D., Bornkamm, G. W. and Polack, A. (1995). Idetification of two enhancer elements downstream of the human *c-myc* gene. Nucleic Acides Res. 23, 72-80

148. Kharroubi, A. E. and Verdin, E. (1994). Protein-DNA interactions within DNase I hypersensitive sites located downstream of HIV-1 promoter. J. Biol. Chem. 269, 19916-19924

149. Fujita, T., Ohtani-Fujita, N. and Sakai, T. (1996). Identification of an RB-responsive region in the 5' untranslated region of the *RB* gene. Cancer Lett. 101, 149-157

150. Hiller, S., Breit, S., Wang, Z, Q., Wagner, E.F. and Schwab, M. (1991). Localization of regulatory elements controlling human *MYCN* expression. Oncogene 6, 969-977

151. Stanton, B. R., Perkins, A. S., Tessarollo, L., Sassoon, D. A. and Parada, L. F. (1992). Loss of *N-myc* function results in embryonic lethality and failure of the epithelial component of the embryo to develop. Genes Dev. 6, 2235-2247

152. Charron, J., Malynn, B. A., Fisher, P., Stewart, V., Jeannotte, L., Goff, S. P., Robertson, E. J. and Alt, F. W. (1992). Embryonic lethality in mice homozygous for a targeted disruption of the *N-myc* gene. Genes Dev. 6, 2248-2257

153. Wedel, A. and Löms Ziegler-Heitbrock, H. W. (1995). The C/EBP family of transcription factors. Immunobiol. 193, 171-185

154. Birkenmeier, E. H., Gwynn, B., Howard, S., Jerry, J., Gordon, J. I., Landschulz. W. H. and Mcknight, S. L. (1989). Tissue-specific expression, developmental regulation, and genetic mapping of the gene encoding CCAAT/enhancer binding protein. Genes Dev. 3, 1146-1156

155. Lin, F. T. and Lane, M. D. (1994). CCAAT/enhancer binding protein α is sufficient to initiate the 3T3-L1 adipocyte differrentiation program. Proc. Natl. Acad. Sci. USA. 91, 8757-8761

156. Yamamoto, M., Ko, L. J., Leonard, M. W., Beug, H., Orkin, S. H. and Engel, J. D. (1990). Activity and tissue-specific expression of the transcription factor NF-E1 multigene family. Genes Dev. 4, 1650-1662

157. Orkin, S. H. (1990). Globin gene regulation and switching: Circa 1990. Cell 63, 665-672

158. Mignotte, V., Wall, L., deBoer, E., Grosveld, F. and Romeo, P. H. (1989). Two tissue-specific factors bind the erythroid promoter of the human porphobilinogen deaminase gene. Nucleic Acids Res. 17, 37-54

159. Brady, H. J. M., Sowden, J. C., Edwards. M., Lowe, N. and Butterworth, P. H. W. (1989). Multiple CF-1 binding sites flank the erythroid specific transcription unit of the human carbonic anhydrase I gene. FEBS Lett. 257, 451-456

160. Pevny, L., Simon, M. C., Robertson, E., Klein, W. H., Tsai, S. F., D' Agati, V., Orkin, S. H. and Costantini, F. (1991). Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. Nature 349, 257-260

161. Ko, L. J., Yamamoto, M., Leonard, M. W., George, K. M., Ting, P. and Engel, J. D. (1991). Murine and human T-Lymophocyte GATA-3 factors mediate transcription through a *cis*-regulatory element within the human T-cell receptor δ gene enhancer. Mol. Cell. Biol. 11, 2778-2784

162. Buskin, J. N. and Hauschka, S. D. (1989). Identification of a myocyte nuclear factor that binds to the muscle-specific enhancer of the mouse muscle creatine kinase gene. Mol. Cell. Biol. 9, 2627-2640

163. Church, G. M., Ephrussi, A., Gilbert, W. and Tonegawa, S. (1985) Cell-type-specific contacts to immunoglobulin enhancers in nuclei. Nature 313, 798-801

164. Lenardo, M., Pierce, J. W. and Baltimore, D. (1988). Protein binding sites in Ig enhancers determine transcriptional activity and inducibility. Science 236, 1573-1577

165. Lassar, A. B., Davis, R. L., Wright, W. E., Kadesch, T., Murre, C., Voronova, A., Baltimore, D. and Weintraub, H. (1991). Functional activity of myogenic HLH proteins requires hetero-oligomerization with E12/E47-like proteins *in vivo*. Cell 66, 305-315

166. Jen, Y., Weintraub, H. and Benezra, R. (1992). Overexpression of Id protein inhibits the muscle differentiation program: *in vivo* association of Id with E2A proteins. Genes Dev. 6, 1466-1479

167. Riggs, K. J., Merrell, K. T., Wilson, G. and Calame, K. (1991). Common factor 1 is a transcriptional activator which binds in the *C-myc* promoter, the skeletal α-actin promoter, and the immnuoglobulin heavy-chain enhancer. Mol. Cell. Biol. 11, 1765-1769

168. Chow, K. L. and Schwartz, R. J. (1990) A combination of closely associated positive and negative *cis*-acting promoter elements regulates transcription of the skeletal α-actin gene. Mol. Cell. Biol. 10, 528-538

169. Perez-Mutul, J., Macchi, M. and Wasylyk, B. (1988). Mutational analysis of the contribution of sequence motifs within the IgH enhancer to tissue specific transcriptional activiation. Nucleic Acids Res. 16, 6085-6096

170. Hernandez, N. (1993). TBP, a universal eukaryotic transcription factor. Genes Dev. 7, 1291-1308

171. Akira, S., Isshiki, H., Sugita, T., Tanabe, O., Kinoshita, S., Nishio, Y., Nakajima, T., Hirano, T. and Kishimoto, T. (1990). A nuclear factor for IL-6 expression (NF-IL6) is a member of a C/EBP family. EMBO J. 9, 1897-1906

172. Hensel, G., Meichle, A., Hizenmaier, K. and Krönke, M. (1989). PMA-responsive 5' flanking sequences of the human TNF gene. Lymphokine Res. 8, 347-351